A thousand empirical adaptive landscapes and their navigability

Aguilar-Rodríguez, José ; Payne, Joshua L ; Wagner, Andreas

Abstract: The adaptive landscape is an iconic metaphor that pervades evolutionary biology. It was mostly applied in theoretical models until recent years, when empirical data began to allow partial landscape reconstructions. Here, we exhaustively analyse 1,137 complete landscapes from 129 eukaryotic species, each describing the binding affinity of a transcription factor to all possible short DNA sequences. We find that the navigability of these landscapes through single mutations is intermediate to that of additive and shuffled null models, suggesting that binding affinity—and thereby gene expression—is readily fine-tuned via mutations in transcription factor binding sites. The landscapes have few peaks that vary in their accessibility and in the number of sequences they contain. Binding sites in the mouse genome are enriched in sequences found in the peaks of especially navigable landscapes and the genetic diversity of binding sites in yeast increases with the number of sequences in a peak. Our findings suggest that landscape navigability may have contributed to the enormous success of transcriptional regulation as a source of evolutionary adaptations and innovations. An adaptive landscape is a mapping from a high-dimensional space of genotypes onto fitness or some other related quantitative phenotype, which defines the 'elevation' of each coordinate in genotype space 1 . Evolution can be viewed as a hill-climbing process in an adaptive landscape, where populations tend to move towards peaks as a consequence of natural selection. The ruggedness of an adaptive landscape has important evolutionary consequences, particularly for the evolution of sex, reproductive isolation and mutational robustness, and for the predictability of evolution 2 . An adaptive landscape that is smooth and single peaked does not pose any obstacle to evolutionary exploration. It is therefore highly navigable, in that it is possible to reach the global peak via positive selection through a series of small mutations that only move 'uphill'. In contrast, a rugged landscape can block the approach to the highest peak by entrapping populations on local suboptimal peaks 3 . We know very little about the navigability of empirical adaptive landscapes, largely due to the incompleteness of the landscapes that have been constructed to date. With few exceptions 4,5 , these landscapes were built by assaying the phenotypes of only a small number of mutations in all possible combinations within a single wild-type background 2 . These studies have helped form our intuition about the structure and navigability of empirical adaptive landscapes, but their conclusions are limited by the fact that they describe only a minute fraction of any complete landscape. An additional caveat of earlier studies is their focus on just one or a few landscapes, which limits the generality of their findings. To study the navigability of a large number of complete, empirical adaptive landscapes, we consider data that describe the binding affinity of a transcription factor (TF)—a sequence-specific DNA-binding protein that helps regulate gene expression—to all possible DNA sequences (TF binding sites) of eight nucleotides in length. TFs are fundamental mediators of gene expression and are involved in numerous evolutionary innovations 6 . Their regulatory effect can be modulated via mutations in TF binding sites, which may alter a TF's affinity for a site and thereby affect gene expression 7,8,9 . We describe the mapping of DNA sequence to binding affinity as an adaptive landscape, where we can study selection for TF binding. This is a common approach for exploring the evolution of TF binding sites 10,11,12,13,14 , other protein–DNA interactions 4,15,16 and protein–RNA interactions 17 . In this context, adaptive evolution is an exploration of sequence space that attempts to optimize the capacity of a sequence to bind a particular TF.

DOI: https://doi.org/10.1038/s41559-016-0045

A thousand empirical adaptive landscapes and their navigability

José Aguilar-Rodríguez[1,2†], Joshua L. Payne[1,2†], Andreas Wagner[1,2,3*]

[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

[2]Swiss Institute of Bioinformatics, Quartier Sorge - Bâtiment Génopode, 1015 Lausanne, Switzerland.

[3]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA.

[†]These authors contributed equally to this work.

*Correspondence: Andreas Wagner
University of Zurich
Winterhurerstrasse 190
Building Y27-J-54
+41-44-635-6141
andreas.wagner@ieu.uzh.ch

Abstract

The adaptive landscape is an iconic metaphor that pervades evolutionary biology. It has been applied mostly in theoretical models until recent years, when empirical data began to allow partial landscape reconstructions. Here, we exhaustively analyze 1,137 complete landscapes from 129 eukaryotic species, each describing the binding affinity of a transcription factor to all possible short DNA sequences. We find that the navigability of these landscapes through single mutations is intermediate to that of additive and shuffled null models, suggesting that binding affinity – and thereby gene expression – is readily fine-tuned via mutations in transcription factor binding sites. The landscapes have few peaks that vary in their accessibility and in the number of sequences they contain. Binding sites in the mouse genome are enriched in sequences found in the peaks of especially navigable landscapes, and the genetic diversity of binding sites in yeast increases with the number of sequences in a peak. Our findings suggest that landscape navigability may have contributed to the enormous success of transcriptional regulation as a source of evolutionary adaptations and innovations.

An adaptive landscape is a mapping from a high-dimensional space of genotypes onto fitness or some other related quantitative phenotype, which defines the "elevation" of each coordinate in genotype space[1]. Evolution can be viewed as a hill-climbing process in an adaptive landscape, where populations tend to move toward peaks as a consequence of natural selection. The ruggedness of an adaptive landscape has important evolutionary consequences, particularly for the evolution of sex, reproductive isolation, mutational robustness, and for the predictability of evolution[2,3]. An adaptive landscape that is smooth and single-peaked does not pose any obstacle to evolutionary exploration. It is therefore highly navigable, in that it is possible to reach the global peak via positive selection through a series of small mutations that only move "uphill." In contrast, a rugged landscape can block the approach to the highest peak by entrapping populations on local suboptimal peaks[4].

We know very little about the navigability of empirical adaptive landscapes, due largely to the incompleteness of the landscapes that have been constructed to date. With few exceptions[5,6], these landscapes were built by assaying the phenotypes of only a small number of mutations in all possible combinations within a single wild-type background[2,3]. These studies have helped form our intuition about the structure and navigability of empirical adaptive landscapes, but their conclusions are limited by the fact that they describe only a minute fraction of any complete landscape. An additional caveat of earlier studies is their focus on just one or a few landscapes, which limits the generality of their findings.

To study the navigability of a large number of complete, empirical adaptive landscapes, we consider data that describe the binding affinity of a TF – a sequence-specific DNA-binding protein that helps regulate gene expression – to all possible DNA sequences (TF binding sites) of eight nucleotides in length. TFs are fundamental mediators of gene expression and are involved in numerous evolutionary innovations[7,8]. Their regulatory effect can be modulated via mutations in TF binding sites, which may alter a TF's affinity for a site and thereby affect gene expression[9–11]. We describe the mapping of DNA sequence to binding affinity as an adaptive landscape, where we can study selection for TF binding. This is a common approach for exploring the evolution of TF binding sites[12–16], other protein-DNA interactions[5,17,18], and protein-RNA interactions[19]. In this context, adaptive evolution is an exploration of sequence space that attempts to optimize the capacity of a

sequence to bind a particular TF.

Results

Adaptive landscapes of transcription factor binding affinity

We obtained protein-binding microarray (PBM) data for 1,137 TFs from the UniPROBE[20] and CIS-BP[21] databases. These TFs represent 129 eukaryotic species and 62 different DNA binding domain structural classes. For each TF, we constructed an adaptive landscape from the enrichment score (*E*-score) – a proxy for relative binding affinity[22–24] – of each of the 32,896 possible sites that bind the TF (Materials and Methods). These landscapes are complete because they describe the affinity with which a TF binds all possible sites, in the absence of confounding factors such as epigenetic marks, chromatin context, local sequence context, or interactions with protein partners. We consider a sequence as "bound" if its *E*-score exceeds 0.35[22,24–26] (Materials and Methods). We use this affinity threshold τ to differentiate sequences that are specifically bound by a TF via hydrogen bond donors and acceptors from those that are unspecifically bound by a TF, e.g., via its affinity for the DNA backbone. To facilitate the analysis of these landscapes, we represent each of them as a genotype network[27], in which vertices represent bound DNA sequences and edges connect sequences that differ by a single point mutation or a short insertion/deletion[24] (Fig. 1A, Materials and Methods). These networks sometimes comprise multiple disconnected components (Supplementary section 3.1, Supplementary Figs. 1-3); when this occurs, we consider only the largest component, which we refer to as the *dominant* genotype network. Each dominant genotype network forms the basis of an adaptive landscape, in which binding affinity defines the "elevation" of each coordinate (TF binding site) in genotype space.

We define landscape navigability as the ability to access a global peak via an evolutionary exploration involving random mutation and natural selection. Landscape navigability is highest when all mutational paths to the global peak exhibit a monotonic increase in binding affinity, which implies a landscape that is smooth and single-peaked. Landscape navigability is lowest when no mutational paths to the global peak exhibit a monotonic increase in binding affinity. This implies a rugged landscape with many peaks. Our measures of landscape navigability depend on two parameters δ and

τ, which we use for noise filtering and for delineating bound from unbound sequences, respectively (Fig. **1**B,C; Materials and Methods). We compare the navigability of the 1,137 empirical landscapes to landscapes generated via two different null models that provide lower and upper bounds on navigability (Materials and Methods). In the first null model (the *additive* model), we deterministically assign a binding affinity to each of the TF binding sites in the genotype network using the TF's position weight matrix, which assumes additive interactions between nucleotides and therefore produces smooth and highly navigable landscapes. In the second null model (the *shuffled* model), we randomly permute the affinities of the TF binding sites in the genotype network, which yields a rugged landscape that hinders navigability. Due to the stochastic nature of the latter model, we generate 1,000 shuffled landscapes for each TF. These null models capture two opposing extremes of landscape navigability while maintaining the structure of the underlying genotype network, thus providing two points of reference for the empirical landscapes.

Landscape navigability: The number of peaks

The number of peaks in an adaptive landscape is an important indicator of its navigability. The more peaks a landscape has, the less navigable it becomes if these peaks are of unequal height. We find that 42% of the empirically derived landscapes (478 of 1,137) have multiple peaks of unequal height (Supplementary Figs. 4 and 5), with peak numbers ranging from 2 to 36 (Fig. 2A). In comparison, only 4.5% of the additive landscapes (51 of 1,137) have multiple peaks, whereas 99% of the shuffled landscapes (1,125,800 of 1,137,000) have multiple peaks (Fig. 2A). One might think that larger landscapes with more binding sites also contain more peaks, and earlier theoretical work hints at that possibility[4]. However, the experimental data we analyse shows no such scaling relationship (Supplementary Fig. 6A). Thus, by the criterion of peak number, smaller landscapes are not necessarily more navigable than larger landscapes. Also, while one usually thinks of a peak as a single sequence, the data do not support this notion. Except for 36 landscapes, the global peaks – those containing the highest-affinity site – are plateaus. They contain between 2 and 121 sequences (Supplementary section 3.2, Supplementary Figs. 5 and 7).

Landscape navigability: Epistasis

Epistasis[28] – non-additive effects of different mutations on a quantitative phenotype (or fitness) – can increase the ruggedness of an adaptive landscape. In the absence of epistasis, adaptive landscapes are smooth and single-peaked, and thus do not hinder evolutionary exploration. Epistasis can be partitioned into three different classes – magnitude, simple sign and reciprocal sign – with increasingly detrimental effects on landscape navigability (Box 1)[29]. To study epistasis, we first identify all squares in a genotype network – a binding site, two of its one-mutant neighbours, as well as the double mutant that can be formed from the single mutants – and classify their affinity relationships according to the scheme in Box 1 (Materials and Methods). In the additive landscapes, on average no more than 0.1% of squares show either of the two kinds of epistasis – simple or reciprocal sign epistasis (here collectively referred to as *sign epistasis*) – that impede landscape navigability most severely (Supplementary section 3.3, Supplementary Fig. 8). In the experimental data, sign epistasis is somewhat more frequent, and affects 4.7% of squares, on average (Fig. 2B). However, its incidence is still 5 times lower than in the shuffled landscapes, where it affects 24.5% of the squares. In the experimental data only magnitude epistasis, which does not affect landscape navigability, approaches the levels observed in the shuffled landscape. We additionally find that sign epistasis preferentially occurs among nearby nucleotides in a binding site[30], whereas magnitude epistasis shows no such preference (Supplementary section 3.4, Supplementary Fig. 9).

Landscape navigability: Accessible mutational paths

Another important indicator of landscape navigability is the fraction of *accessible* mutational paths to a given genotype from all other genotypes in the landscape (Supplementary Fig. 10). Here, a mutational path is considered accessible if each mutation in the path increases binding affinity monotonically[31]. Figure 2C shows the fraction of accessible paths to the highest-affinity binding site in a global peak (i.e., peak accessibility), in relation to the length of the mutational path, for the empirical data and both null models. In all three cases, the fraction of accessible mutational paths to the highest-affinity site decreases with the length of the path. However, the rate of decrease for the

empirical data is intermediate to that of the additive and shuffled models, indicating that the empirical landscapes are always less navigable than the additive landscapes, but more navigable than the shuffled landscapes. Even for the longest mutational paths to the highest-affinity site, more than 20% of paths are accessible in the empirical landscapes. Moreover, even when unbound sequences are included in the landscapes – a modification that decreases landscape navigability – the global peak remains accessible for all but the longest mutational paths (Supplementary Fig. 11; Supplementary section 3.5).

Taken together, these 3 measures of landscape navigability – number of peaks, epistasis, and peak accessibility – indicate that transcription factor binding affinity landscapes are more navigable than shuffled landscapes, but less navigable than additive landscapes, a conclusion that is robust to broadly varying parameter choices (δ and τ) and modelling assumptions (Supplementary section 3.6; Supplementary Figs. 12-30). While experiments are required to determine which parameters and assumptions best reflect the true binding affinity landscapes, the following *in vivo* analyses suggest that the baseline parameter combination studied here provides meaningful information about TF binding in both yeast and mouse, two highly diverged eukaryotic species.

Navigability influences the *in vivo* abundance of binding sites

Landscape navigability varies among TFs within the same species. This led us to reason that global peak sequences from more navigable landscapes might be more abundant in the regulatory regions of living organisms than global peak sequences from less navigable landscapes. The reason is that smooth landscapes pose fewer obstacles to the evolution of global peak sequences than rugged landscapes. To test this hypothesis, we considered two sources of *in vivo* data from 14 cell and tissue types in *Mus musculus*: RNA-seq transcript abundance estimates[32] and maps of genome-wide DNase I footprints[33]. The RNA-seq data indicate which TFs are expressed in each of the 14 cell and tissue types, which is important because we only expect landscape navigability to impact the *in vivo* abundance of a binding site if its cognate TF is expressed in that cell or tissue type. The DNase I footprints demarcate DNA sequences in open chromatin that are bound by protein[34] (Materials and Methods), and can therefore be used to predict TF binding sites.

For the 187 murine TFs in our dataset that were expressed in a given cell or tissue type (Materials and Methods, Supplementary Table 2), we determined the *in vivo* abundance of the TF's highest-affinity binding site by counting the number of times the site appears in DNase I footprints that are predicted to bind the TF. We determined the statistical significance of each count by comparing it to the number of times the sequence is expected to appear in stretches of DNA that have the same length and the same mono- and di-nucleotide frequencies as the footprints (Materials and Methods). We find that the highest-affinity sites in landscapes with multiple peaks are less abundant in regulatory regions genome-wide than those from landscapes with a single peak across all of the 14 cell and tissue types (Supplementary Fig. 31, Supplementary Table 3), as shown for heart tissue in Fig. 3A (Wilcoxon rank-sum test, *P*-value = 1.42×10^{-6}). We also observe that highly accessible global peak sequences are more abundant in protein-bound regions of the mouse genome than are less accessible global peak sequences across 10 of the 14 cell and tissue types (Fig. 3B; Supplementary Table 3, Supplementary Fig. 32). As a negative control, we repeated the above analyses using the DNase I hypersensitive regions that flank the footprints, rather than the footprints themselves. For this control data, the abundance of the highest-affinity sites is not significantly associated with the number of peaks in any of the 14 cell and tissue types (Supplementary Table 4). Also for the control data, and in 13 of the 14 tissue types, the abundance of the highest-affinity sites is not significantly associated with peak accessibility, and in the remaining tissue, this association is only marginally significant (Heart; Spearman's rank correlation coefficient = 0.27, *P*-value = 0.034; Supplementary Table 4). Importantly, the effects of landscape navigability on binding site abundance still hold after controlling for binding affinity (Supplementary section 3.7, Supplementary Tables 5 and 6), though statistical significance is lost in three tissues for peak accessibility. Moreover, our observations hold in all cell and tissue types after controlling for the information content of each TF's position weight matrix (Supplementary section 3.7, Supplementary Tables 5 and 7). Taken together, these findings suggest that landscape navigability has influenced the evolution of TF binding sites in the mouse genome.

Gene expression reflects landscape topography

Gene expression levels can be fine-tuned via affinity-altering mutations in TF binding sites[9,11]. Models of regulatory evolution commonly assume a direct mapping between binding affinity and gene expression[35,36], such that monotonic changes in binding affinity lead to monotonic changes in gene expression. We tested this assumption with *in vivo* gene expression data from a recent high-throughput promoter screen in *Saccharomyces cerevisae*[11]. These data comprise replicated *in vivo* gene expression measurements for every single-base-pair and many double- and triple-base-pair mutants of a TF's consensus binding site (Materials and Methods). Protein-binding microarray data are available for two of these TFs (Gcn4 and Fhl1), facilitating the superposition of *in vivo* transcriptional output with *in vitro* binding affinity. For the subset of each TF's genotype network where *in vivo* expression data are available, we determined whether gene expression levels increase monotonically along accessible mutational paths to the site with the highest affinity. For Gcn4, all 71 accessible mutational paths exhibit monotonic increases in gene expression (P = 0.01, permutation test; Fig. 4A), and for Fhl1 all except one of the 37 accessible mutational paths exhibit monotonic increases in gene expression (P = 0.001, permutation test; Fig. 4B). These findings suggest that, at least for these two TFs, the navigability of the adaptive landscape of binding affinity facilitates the evolution of increased gene expression.

We also compared the binding affinity landscapes of Gcn4 and Fhl1 to incomplete landscapes constructed using gene expression data from the high-throughput promoter screens. To do so, we developed a measure of similarity between the complete *in vitro* landscapes and the incomplete *in vivo* landscapes. This measure is simply the sum of the absolute differences between binding affinity and gene expression for each of the binding sites. We reasoned that if the *in vitro* binding affinity landscapes are truly informative of *in vivo* gene expression, then this sum would be significantly smaller than expected if instead the gene expression measurements were randomly permuted. To test this hypothesis, we compared the observed sum to a null distribution of sums obtained via 10^5 random permutations of the gene expression data. The fraction of permuted data sets in which the sum is smaller than that of the measured expression data yields an empirical P-value for each TF. Based on this test, *in vitro* binding affinity is indeed informative of *in vivo* gene expression (Gcn4: P-value =

0.0011; Fhl1: *P*-value = 0.0125; Fig. 4C,D), which provides additional validation of the binding affinity landscapes studied here.

Global peak breadth affects the diversity of binding sites

We have shown that global peaks typically comprise many different binding sites of similar affinity. The broader a peak is, the more of its sequences have mutant neighbours with approximately the same affinity. One would thus expect that sequences in broader peaks could evolve more freely by means of nucleotide changes and thus accumulate greater genetic diversity. In contrast, mutations in the binding sites of narrower peaks – those with fewer binding sites – will more often lead to a decrease in binding affinity and thus be eliminated. To find out if global peak breadth has any effect on binding site evolution, we asked whether global peak breadth affects genetic diversity in binding sites. Using single nucleotide polymorphism data across 19 strains of *S. cerevisiae*[37], we calculated the diversity within binding sites[38] of 23 different TFs as the average Shannon's diversity index per site (Materials and Methods). Indeed, the broader the global peak of a TF's affinity landscape, the greater is the average diversity of polymorphic sites that bind the TF (Spearman's rank correlation coefficient = 0.73, *P*-value = 8.9×10^{-5}; Fig. 5). We emphasize that this trend is not driven by TF specificity, since the information content of the TFs' position weight matrices exhibit no correlation with binding site diversity (Spearman's rank correlation coefficient = -0.08, *P*-value = 0.71). This finding provides further validation that the topography of a binding affinity landscape can impact the evolution of TF binding sites.

Discussion

We have used measurements from protein-binding microarrays to construct and analyse more than one thousand complete, empirical adaptive landscapes, each describing the binding affinity of a TF to all possible short DNA sequences. Such landscapes are important objects of study, because changes in the level, location, or timing of gene expression commonly underlie evolutionary innovations[39–41], and gene expression patterns are readily fine-tuned via small changes in binding affinity[9,11]. Understanding how an evolutionary process might navigate a TF binding affinity landscape is therefore an important step toward understanding how gene regulatory programs evolve. Here, we

have taken this step, demonstrating that the navigability of binding affinity landscapes is intermediate to that of additive and shuffled (rugged) landscapes, but closer to the additive expectation in terms of the number of peaks and the incidence of 3 forms of epistasis.

Our measures of landscape navigability allow us to understand how individual binding sites can evolve toward higher binding affinity in the absence of confounding factors (such as chromatin context[42]), a pursuit that is motivated by earlier theoretical work[12,15,35] and by several empirical observations which indicate that high-affinity sites are used preferentially *in vivo*. For example, studies of blastoderm patterning in *Drosophila melanogaster* have shown that high-affinity sites typically reside near probable functional targets, whereas low-affinity sites are more often found near genes that are not transcribed in the early embryo[43], and that are less likely to drive expression in transgenic reporter assays[44]. Moreover, in both microbes and humans, affinity-decreasing mutations are predominately under negative selection, whereas affinity-increasing mutations are under positive selection[45,46], consistent with the inferred monotonic increases in organismal fitness that accompany increased binding affinity[15,16,47]. Nevertheless, low-affinity sites do sometimes play important regulatory roles[48–51] and it is therefore worth noting that, by symmetry, accessible paths to the highest-affinity site are also accessible in the opposite direction by permitting only monotonically decreasing changes in binding affinity. More generally, the landscapes constructed here can easily be transformed to study selection for low or intermediate binding affinity, by assuming that the fitness conferred by a binding site is a decreasing function of the difference between the site's affinity and an arbitrary optimal affinity[35]. Understanding how such transformations affect the navigability of TF binding affinity landscapes is an exciting direction for future work.

To summarize, while our analyses of *in vitro* and *in vivo* measurements of TF-DNA interactions have some caveats (Supplementary section 4), they suggest that the navigability of transcription factor binding affinity landscapes has left a trace in the mouse and yeast genomes' portfolio of regulatory DNA. Landscape navigability may therefore have contributed to the enormous success of transcriptional regulation as an evolutionary mechanism for generating variation and innovation.

Materials and Methods

In vitro data

The *in vitro* data we have studied come from protein-binding microarrays (PBM)[23,52], which measure the binding affinity of a TF to all 32,896 possible eight-nucleotide, double-stranded DNA sequences. There are $(4^8 - 4^4) / 2 + 4^4 = 32{,}896$ sequences, rather than $4^8 = 65{,}536$ sequences, because each sequence is merged with its reverse complement and because there are 4^4 sequences that are identical to their reverse complement and therefore cannot be merged.

We had three criteria for including a TF in our dataset. First, it had to be analysed on two different PBM designs. Second, it had to bind at least one DNA sequence with an *E*-score above 0.45, as this indicates a high level of data quality[21]. Third, its genotype network (Fig. 1A) had to contain at least one square to permit the analysis of epistasis (Box 1). Based on these criteria, we obtained PBM data for 42 yeast[25] and 104 mouse TFs[22] from the UniPROBE database[20] and for 991 TFs belonging to 129 different eukaryotic species (including 4 additional TFs for *S. cerevisiae* and 83 for *M. musculus*) from the CIS-BP database[21]. In total, our dataset comprises 1,137 TFs, representing 129 eukaryotic species and 62 DNA binding domain structural classes (Supplementary Table 1).

For each TF, the PBM data include a non-parametric, rank-based enrichment score (*E*-score) for each of the 32,896 DNA sequences. The *E*-score is a variant of the Wilcoxon-Mann-Whitney statistic[23] and ranges from −0.5 (most disfavoured site) to 0.5 (most favoured site). *E*-scores correlate with the relative dissociation constants of TFs[22,23] and can be used as a proxy for relative binding affinity. We therefore refer to this measure as binding affinity and use it to delineate bound from unbound sequences, and as the quantitative phenotype that defines the surface of our adaptive landscapes. Except for the 42 yeast TFs from Zhu *et al.*[25], the data also includes a median signal intensity *Z*-score for each site. We use this score as an alternative proxy for binding affinity in Supplementary section 3.6.5. Following earlier work[24,25], we only consider a sequence as bound by a TF if its *E*-score exceeds a threshold of 0.35. The reason for this threshold is that it has precedent[22,26] and, more importantly, an analysis of the relationship between *E*-score and false discovery rate (FDR) in 104 mouse TFs[22] revealed that all sequences with an *E*-score exceeding 0.35 had an FDR below

0.001. This threshold can therefore be used to delineate sequences that are specifically bound by a TF from those that are unspecifically bound. In addition to the threshold $\tau = 0.35$, we conducted a sensitivity analysis of our results by broadly varying τ (see Supplementary section 3.6.2).

***In vivo* data**

We collected DNase I footprints for 14 cell and tissue types in *M. musculus*[33]. These genome-wide data specify DNA sequences that are in open chromatin and bound by protein, at single nucleotide resolution. For each of the 187 murine TFs in our dataset, we used FIMO[53] and the TF's position weight matrix (obtained from UniPROBE[20] and CIS-BP[21]) to scan these footprints for potential binding sites using a $P < 1\times10^{-4}$ threshold. We then counted the number of times that each eight-nucleotide DNA sequence appeared in the predicted binding sites, considering both strands of the DNA. To determine the statistical significance of each count, we compared it to the number of times the same sequence is expected to appear in stretches of DNA that have the same length and mono- and di-nucleotide frequencies as the footprints of each cell and tissue type, thus controlling for the GC content of the footprints. Following van Helden *et al.*[54], the statistical significance of each observed count was determined using the binomial formula, with a conservative significance threshold of $P < 1/32896$ – the inverse of the number of possible DNA sequences of length eight. Across the 14 cell and tissue types, the counts for the highest-affinity sites of each of the 187 mouse TFs in our dataset ranged from zero to 73,174.

We also collected the DNase I hypersensitive regions that flank the footprints in each of the 14 cell and tissue types. We used these regions to perform a negative control, in which we correlated our landscape navigability measures with *in vivo* binding site abundance within regions of open chromatin that do not show evidence of protein binding. Specifically, we counted the number of times that each eight-nucleotide DNA sequence appeared in the DNase I hypersensitive regions, after having removed the footprints of the 14 cell and tissue types. To determine the statistical significance of each count, we compared it to the number of times the same sequence is expected to appear in stretches of DNA that have the same length and mono- and di-nucleotide frequencies as the DNase I hypersensitive regions of each cell and tissue type, again using the binomial formula and a stringent significance threshold ($P < 1/32896$).

To determine which of the 187 TFs are expressed in each of the 14 cell and tissue types, we collected RNA-seq data for the same cell and tissue types[32] and used cufflinks[55] to calculate the FPKM (fragments per kilobase of transcript per million fragments mapped) for each TF. We considered a TF as expressed in a given cell or tissue type if its FPKM ≥ 1. Supplementary Table 2 shows that the number of TFs expressed in a given cell or tissue type ranges from 89 to 133. This table also provides the file names of the DNase I hypersensitive regions, DNase I footprints, and the RNA-seq data that we analysed.

For two yeast TFs (Gcn4 and Fhl1), we also collected gene expression data from high-throughput promoter screens[11]. These data include gene expression measurements from a large library of engineered promoters, each 150 nucleotides long. Each of these promoters contains between zero and three point mutations to the TF's consensus sequence. For Gcn4, which binds a seven-nucleotide sequence, this library includes the consensus sequence (TGACTCA), all 21 single-mutants, 42 double-mutants, and 10 triple-mutants (73 sequences in total). For Fhl1, which binds an eight-nucleotide sequence, this library includes the consensus sequence (GACGCAAA), all 24 single-mutants, 56 double-mutants, and 10 triple-mutants (90 sequences in total). For each promoter in the library, the data include gene expression measurements from two biological replicates. We used the average gene expression measurement per promoter.

To map the gene expression measurements for the seven-nucleotide sequences bound by Gcn4 onto the eight-nucleotide sequences for which we have protein-binding microarray data, we first located each of the 73 seven-nucleotide sequences within the 150-nucleotide promoters. We then padded each of the seven-nucleotide sequences with one nucleotide upstream and one nucleotide downstream from its respective promoter, forming two eight-nucleotide sequences. We assigned the gene expression measurement for the seven-nucleotide sequence to each of the eight-nucleotide sequences that were formed in this fashion. This procedure generated gene expression measurements for 132 eight-nucleotide sequences.

For the subset of the sequences for which we have measurements of both *in vitro* binding affinity and *in vivo* gene expression, we determined the number of accessible mutational paths to the sequence with the highest level of binding affinity. For Gcn4, there are 71 accessible mutational paths: 30 of

length 3, 37 of length 2 and 14 of length 1. For Fhl1, there are 37 accessible mutational paths: 22 of length 2 and 15 of length 1. For each TF, we then determined the fraction of these accessible paths in which gene expression increased monotonically. To determine whether two gene expression levels truly differed from one another, we used a noise threshold that covered the same proportional range of expression levels as covered by the noise threshold δ in the range of affinity values. To calculate the statistical significance of the fraction of accessible paths in which gene expression increased monotonically, we performed a permutation test, in which we randomly permuted the sequences' gene expression measurements, while preserving their binding affinities. We repeated this process 1000 times for both Gcn4 and Fhl1.

We obtained the genomic coordinates of TF binding sites in *S. cerevisae* from a map of conserved regulatory sites[38], and analysed binding sites detected with a stringent binding *P*-value cut-off of 0.001, but no conservation cut-off. We obtained single nucleotides polymorphism (SNP) data for nineteen different strains from the Saccharomyces Genome Resequencing Project (SGRP)[37]. The TF binding sites were based on the January 2006 Saccharomyces Genome Database (SGD) version of the reference strain *S. cerevisae* S288c genome sequence, while the SNP data were based on the January 2010 version. We used liftOver to convert both sets of genomic coordinates to the coordinates in the February 2011 version. Then we used the *intersect* function from the BEDTools suite (version 2.25.0)[56] to determine the presence of SNPs within binding sites. Using these data, we calculated the genetic diversity of polymorphic TF binding sites across the nineteen yeast strains. For each position in a binding site we calculated *Shannon's diversity index* (H):

$$H = -\sum_i p_i \ln(p_i) \quad \textbf{(1)}$$

where p_i is the frequency of allele i, which we computed as the fraction of strains with nucleotide i. We computed the diversity (D) of a binding site as the average of H over all L positions:

$$D = \frac{1}{L}\sum_{j=1}^{L} H_j \quad \textbf{(2)}$$

Genotype networks

The procedure for constructing genotype networks of TF binding sites has been described elsewhere[24]. In brief, for each TF, we first determined the set of sequences that were bound by the TF (*E*-score > 0.35). We then used an alignment algorithm to calculate the mutational distance between all pairs of bound sequences. Finally, we used these mutational distances to define the edges of the genotype network, connecting two sequences if they differed by a single small mutation. The mutations we considered are point mutations and small indels that shift an entire contiguous binding site by a single base[24].

Quantitative measures of landscape navigability

We used several measures to quantify the navigability of an adaptive landscape. All of them are parameterized by δ, a threshold value that is used to determine whether two affinity values truly differ from one another. This parameter is necessary because PBM data are inherently noisy[52]. For each TF, we calculate δ as the residual standard error of a linear regression between the affinity values of all bound sequences from the two replicate PBMs. Thus, each TF has its own δ, which reflects the noise in the replicated PBM measurements for that particular TF (Supplementary Table 1). We say that the binding affinity E_i of site i is greater than the binding affinity E_j of site j if $E_i > E_j + \delta$. Analogously, we say that the binding affinity E_i of site i is less than the binding affinity E_j of site j if $E_i + \delta < E_j$. Otherwise, we say that we cannot differentiate between the two affinity values. The average empirical value of δ across the 1,137 TFs is 0.028, which covers 18.4% of the range of affinity values for bound sequences ($0.35 < E\text{-score} \leq 0.50$). Our criterion for considering two affinity values as different is therefore highly conservative. In addition to these empirical values of δ, we conducted a sensitivity analysis of our results by broadly varying δ (0, 0.001, 0.01, 0.03 and 0.05), above and below the average empirical value (see Supplementary section 3.6.1).

Our first measure of the navigability of an adaptive landscape is its *number of peaks*. To detect a peak, we follow a procedure similar to one previously described[57]. We categorize each sequence in a genotype network as belonging to a peak (either as an individual sequence or as a member of a plateau) or not. To do so, we select sequences in decreasing order of binding affinity to seed a

breadth-first search of the genotype network. Each iteration of the search considers sequences that are an additional mutational step away from the seed sequence. In the first iteration of the search, we determine whether the seed sequence is a peak or not: It is a peak if all of its neighbours in the genotype network have lower binding affinity. It is not a peak if at least one of its neighbours has higher binding affinity. If the seed sequence has at least one neighbour with an affinity that is neither greater nor lower, then these neighbours are retained as belonging to a plateau that may be a peak, and the breadth-first search continues. If at any subsequent iteration of the search a sequence is found that neighbours any of those on the plateau and has an affinity that is higher than the seed sequence, then the seed sequence does not belong to a peak and the search halts. If at any iteration of the search, all sequences that neighbour those on the plateau have lower affinity than the seed sequence, then the seed sequence belongs to a plateau that is a peak.

Our second measure of landscape navigability is based on the concept of accessible mutational paths[31,58,59]. A mutational path is the shortest path that connects two bound sequences *i* and *j* on a genotype network, such that sequence *i* can be transformed into sequence *j* via a series of intermediates that are also on the genotype network. A mutational path from sequence *i* to sequence *j* is considered accessible if binding affinity increases monotonically along the path. We report the fraction of mutational paths that are accessible, starting from all sequences in the genotype network and ending at the highest-affinity site in the global peak. We refer to this fraction as *peak accessibility*. We note that, by symmetry, accessible paths to the highest-affinity site are also accessible in the opposite direction by permitting only monotonically decreasing changes in binding affinity. Accessibility of low-affinity binding sites is important, as such sites are known to play crucial roles in the regulation of certain genes[49,51,60].

Our final measures of landscape navigability pertain to *epistasis*, motivated by a recent debate over the significance of non-additive interactions between the individual bases of TF binding sites in their contribution to binding affinity[61,62]. Epistasis can have detrimental effects on landscape navigability. We quantify the incidence of three classes of epistasis for each of the 1,137 adaptive landscapes[29]. First, we detect all squares in each TF's genotype network. A square is a quadruplet of sequences that contain a binding site, two of its one-mutant neighbours, and the double mutant that

can be formed from the single mutants (Box 1). Second, we designate the highest-affinity sequence as the "double mutant," thus forcing the labelling of the other three sites as the wild-type or the single-mutants. Third, we calculate the magnitude of epistasis as:

$$\varepsilon = E_{AB} + E_{ab} - E_{Ab} - E_{aB} \quad \textbf{(3)}$$

where E_{AB} is the binding affinity of the double mutant, E_{ab} is the binding affinity of the wild type and E_{Ab} and E_{aB} are the binding affinities of the single-mutants. We only consider a mutational pair to be epistatic if $|\varepsilon|$ is greater than or equal to the noise threshold δ.

If this condition is met, we classify the epistatic interaction as magnitude epistasis, simple sign epistasis, or reciprocal sign epistasis[29]. *Magnitude epistasis* occurs when the following relation holds:

$$|\Delta E_{ab \to Ab} + \Delta E_{aB \to AB}| = |\Delta E_{ab \to Ab}| + |\Delta E_{aB \to AB}| \quad \textbf{(4)}$$

where ΔE is the *mutational effect*, i.e., the change in binding affinity caused by a mutation (e.g., $ab \to Ab$). *Simple sign epistasis* occurs instead when the following relation holds:

$$|\Delta E_{ab \to Ab} + \Delta E_{aB \to AB}| < |\Delta E_{ab \to Ab}| + |\Delta E_{aB \to AB}| \quad \textbf{(5)}$$

Reciprocal sign epistasis occurs when both **[3]** and the following relation holds:

$$|\Delta E_{ab \to aB} + \Delta E_{Ab \to AB}| < |\Delta E_{ab \to aB}| + |\Delta E_{Ab \to AB}| \quad \textbf{(6)}$$

If a mutational effect is smaller than δ, we assigned it a value of zero. If all mutational effects are smaller than δ, even despite $|\varepsilon| \geq \delta$ being true, then we classify the interaction as non-epistatic. Taken together with the fact that we excluded unbound sequences from these calculations, our measures of epistasis are conservative (Supplementary section 3.8, Supplementary Fig. 33).

Null models

For each of the 1,137 TFs, we consider two null models. Both change the topography of the adaptive landscape, but maintain the structure of the underlying genotype network. The first null model (the *additive* model) is based on a TF's position weight matrix (PWM)[63]. A PWM represents the binding

preferences of a TF as a $4 \times L$ matrix, where each row corresponds to one of the four bases and each column corresponds to one of the L positions in the binding site. Each matrix entry $f_{i,b}$ is the frequency of base b at position i. We obtained the PWMs of the 1,137 TFs from the UniPROBE[20] and CIS-BP[21] databases.

Since the width L of a TF's PWM may not equal eight, we use a sliding-window to assign a score (S_{pwm}) to each of the eight-nucleotide sequences in each genotype network. Specifically, we slide each sequence on a genotype network through the corresponding TF's PWM from left to right, assigning a score to each of the subsequences. The sliding-window procedure is carried out in such a way that the first scored subsequence is just the single right-most position of the sliding sequence and occupies the left-most column of the PWM. The procedure ends with a subsequence that just corresponds to the single left-most position in the sequence and occupies the right-most column of the PWM. We then repeat this process with the sequence's reverse complement. We take the maximum of these scores as the score for the sequence. As an example, if L is equal to eight, then we take the maximum over 30 separate scores: 15 for the sequence and another 15 for its reverse complement. Each of these scores is calculated as

$$S_{\text{pwm}} = \sum_{i=1}^{l} f_{i,b}\, I(i), \tag{7}$$

where l is the window length of the sliding sequence and $I(i)$ is the information content at position i:

$$I(i) = 2 + \sum_{b} f_{i,b} \log_2 f_{i,b} \tag{8}$$

Base matches at positions with high information content thus contribute more to a sequence's overall score than base matches at positions with low information content. Importantly, this scoring technique is purely additive, i.e., the contribution of each binding site position to the overall score is independent of the other positions in the binding site. To facilitate comparison among TFs, we normalize the scores by the maximum score in each genotype network. To analyse the navigability of

landscapes constructed with this model, we use a noise threshold δ that covers the same range of scores as the δ used for the empirical data.

In the second null model (the *shuffled* model), we randomly permute the binding affinities of the sequences in the genotype network, yielding a rugged landscape topography. Since this process is stochastic, we repeat it 1000 times per TF. In these landscapes, we use the same δ as used for the empirical landscapes, because unlike the additive model, the shuffled null model does not change the range of affinity values.

References

1. Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. in *Proceedings of the Sixth International Congress on Genetics* **1,** 356–366 (1932).

2. Szendro, I. G., Schenk, M. F., Franke, J., Krug, J. & de Visser, J. A. G. M. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech. Theory Exp.* **2013,** P01005 (2013).

3. de Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15,** 480–490 (2014).

4. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128,** 11–45 (1987).

5. Rowe, W. *et al.* Analysis of a complete DNA-protein affinity landscape. *J. R. Soc. Interface* **7,** 397–408 (2010).

6. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 14984–9 (2013).

7. Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8,** 206–16 (2007).

8. Prud'homme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* **104 Suppl,** 8605–12 (2007).

9. Gertz, J., Siggia, E. D. & Cohen, B. A. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457,** 215–8 (2009).

10. Shultzaberger, R. K., Malashock, D. S., Kirsch, J. F. & Eisen, M. B. The fitness landscapes of cis-acting binding sites in different promoter and environmental contexts. *PLoS Genet.* **6,** e1001042 (2010).

11. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30,** 521–30 (2012).

12. Gerland, U. & Hwa, T. On the selection and evolution of regulatory DNA motifs. *J. Mol. Evol.* **55,** 386–400 (2002).

13. Berg, J., Willmann, S. & Lässig, M. Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.* **4,** 42 (2004).

14. Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315,** 233–7 (2007).

15. Mustonen, V., Kinney, J., Callan, C. G. & Lässig, M. Energy-dependent fitness: a quantitative model for the evolution of yeast transcription factor binding sites. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 12376–81 (2008).

16. Haldane, A., Manhart, M. & Morozov, A. V. Biophysical fitness landscapes for transcription factor binding sites. *PLoS Comput. Biol.* **10,** e1003683 (2014).

17. Carlson, C. D. *et al.* Specificity landscapes of DNA binding molecules elucidate biological function. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 4544–9 (2010).

18. Weghorn, D. & Lässig, M. Fitness landscape for nucleosome positioning. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10988–93 (2013).

19. Buenrostro, J. D. *et al.* Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* **32,** 562–8 (2014).

20. Newburger, D. E. & Bulyk, M. L. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **37,** D77-82 (2009).

21. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158,** 1431–43 (2014).

22. Badis, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324,** 1720–3 (2009).

23. Berger, M. F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24,** 1429–35 (2006).

24. Payne, J. L. & Wagner, A. The robustness and evolvability of transcription factor binding sites. *Science* **343,** 875–7 (2014).

25. Zhu, C. *et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res.* **19,** 556–66 (2009).

26. Nakagawa, S., Gisselbrecht, S. S., Rogers, J. M., Hartl, D. L. & Bulyk, M. L. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 12349–54 (2013).

27. Maynard Smith, J. Natural selection and the concept of a protein space. *Nature* **225,** 563–4 (1970).

28. Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27,** 323–31 (2011).

29. Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J. & Tans, S. J. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272,** 141–4 (2011).

30. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152,** 327–39 (2013).

31. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312,** 111–4 (2006).

32. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515,** 355–364 (2014).

33. Stergachis, A. B. *et al.* Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* **515,** 365–370 (2014).

34. Hesselberth, J. R. *et al.* Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat. Methods* **6,** 283–9 (2009).

35. Lynch, M. & Hagner, K. Evolutionary meandering of intermolecular interactions along the drift barrier. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E30-8 (2015).

36. MacArthur, S. & Brookfield, J. F. Y. Expected rates and modes of evolution of enhancer sequences. *Mol. Biol. Evol.* **21,** 1064–1073 (2004).

37. Bergström, A. *et al.* A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* **31,** 872–88 (2014).

38. MacIsaac, K. D. *et al.* An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* **7,** 113 (2006).

39. Keys, D. N. *et al.* Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science* **283,** 532–534 (1999).

40. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. a & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **433,** 481–487 (2005).

41. Rister, J. *et al.* Single-base pair differences in a shared motif determine differential Rhodopsin expression. *Science* **350,** 1258–61 (2015).

42. Siggers, T. & Gordân, R. Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res.* **42,** 2099–111 (2014).

43. Li, X. Y. *et al.* Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol.* **6,** 0365–0388 (2008).

44. Fisher, W. W. *et al.* DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21330–5 (2012).

45. Mustonen, V. & Lässig, M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet.* **25,** 111–119 (2009).

46. Arbiza, L. *et al.* Genome-wide inference of natural selection on human transcription factor binding sites. *Nat. Genet.* **45,** 723–729 (2013).

47. Mustonen, V. & Lässig, M. Evolutionary population genetics of promoters: predicting binding sites and functional phylogenies. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15936–15941 (2005).

48. Swanson, C. I., Schwimmer, D. B. & Barolo, S. Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* **21,** 1186–1196 (2011).

49. Grönlund, A., Lötstedt, P. & Elf, J. Transcription factor binding kinetics constrain noise suppression via negative feedback. *Nat. Commun.* **4,** 1864 (2013).

50. Ramos, A. I. & Barolo, S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20130018 (2013).

51. Crocker, J. *et al.* Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160,** 191–203 (2015).

52. Berger, M. F. & Bulyk, M. L. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nat. Protoc.* **4,** 393–411 (2009).

53. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27,** 1017–1018 (2011).

54. van Helden, J., André, B. & Collado-Vides, J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J. Mol. Biol.* **281,** 827–842 (1998).

55. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).

56. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–2 (2010).

57. Dawid, A., Kiviet, D. J., Kogenaru, M., de Vos, M. & Tans, S. J. Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape. *Chaos* **20,** 26105 (2010).

58. Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445,** 383–6 (2007).

59. Franke, J., Klözer, A., de Visser, J. A. G. M. & Krug, J. Evolutionary accessibility of mutational pathways. *PLoS Comput. Biol.* **7,** e1002134 (2011).

60. Parker, D. S., White, M. A., Ramos, A. I., Cohen, B. A. & Barolo, S. The cis-regulatory logic of Hedgehog gradient responses: key roles for gli binding affinity, competition, and cooperativity. *Sci. Signal.* **4,** ra38 (2011).

61. Zhao, Y. & Stormo, G. D. Quantitative analysis demonstrates most transcription factors require only simple models of specificity. *Nat. Biotechnol.* **29,** 480–3 (2011).

62. Morris, Q., Bulyk, M. L. & Hughes, T. R. Jury remains out on simple models of transcription factor specificity. *Nat. Biotechnol.* **29,** 483–4 (2011).

63. D'haeseleer, P. What are DNA sequence motifs? *Nat. Biotechnol.* **24,** 423–5 (2006).

64. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59,** 1165–74 (2005).

65. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

66. Sokal, R. R. & Rohlf, F. J. *Biometry*. (Freeman, 1995).

67. Provine, W. B. *Sewall Wright and Evolutionary Biology*. (University of Chicago Press, 1986).

68. *The Adaptive Landscape in Evolutionary Biology*. (Oxford University Press, 2012).

69. Hinkley, T. *et al.* A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat. Genet.* **43,** 487–9 (2011).

70. Podgornaia, A. I. & Laub, M. T. Protein evolution. Pervasive degeneracy and epistasis in a protein-protein interface. *Science* **347,** 673–7 (2015).

71. Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352,** 837–40 (2016).

72. Puchta, O. *et al.* Network of epistatic interactions within a yeast snoRNA. *Science* **352,** 840–4 (2016).

73. Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J. & Lehner, B. The complete local genotype-phenotype landscape for the alternative splicing of a human exon. *Nat. Commun.* **7,** 11558 (2016).

74. Salverda, M. L. M. *et al.* Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* **7,** e1001321 (2011).

75. du Plessis, L., Leventhal, G. E. & Bonhoeffer, S. How Good Are Statistical Models at Approximating Complex Fitness Landscapes? *Mol. Biol. Evol.* **33,** 2454–68 (2016).

76. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490,** 535–8 (2012).

77. McCandlish, D. M., Rajon, E., Shah, P., Ding, Y. & Plotkin, J. B. The role of epistasis in protein evolution. *Nature* **497,** E1–E2 (2013).

78. Man, T. K. & Stormo, G. D. Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. *Nucleic Acids Res.* **29,** 2471–2478 (2001).

79. Bulyk, M. L., Johnson, P. L. F. & Church, G. M. Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors. *Nucleic Acids Res.* **30,** 1255–61 (2002).

80. Anderson, D. W., McKeown, A. N. & Thornton, J. W. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *Elife* **4,** 1–26 (2015).

81. Dubertret, B., Liu, S., Ouyang, Q. & Libchaber, A. Dynamics of DNA-protein interaction deduced from in vitro DNA evolution. *Phys. Rev. Lett.* **86,** 6022–6025 (2001).

82. von Hippel, P. H. & Berg, O. G. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 1608–1612 (1986).

83. Berg, O. G., Hippel, P. H. & von Hippel, P. H. Selection of DNA binding sites by regulatory proteins. *Trends Biochem. Sci.* **13,** 207–211 (1988).

84. Sengupta, A. M., Djordjevic, M. & Shraiman, B. I. Specificity and robustness in transcription

control networks. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 2072–2077 (2002).

85. Gerland, U., Moroz, J. D. & Hwa, T. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 12015–12020 (2002).

86. Tuğrul, M., Paixão, T., Barton, N. H. & Tkačik, G. Dynamics of transcription factor binding site evolution. *PLoS Genet.* **11,** e1005639 (2015).

87. Buchler, N. E., Gerland, U. & Hwa, T. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 5136–5141 (2003).

88. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 15690–5 (2015).

89. Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2,** e00726 (2013).

90. Raveh-Sadka, T. *et al.* Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat. Genet.* **44,** 743–750 (2012).

91. Levo, M. *et al.* Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* gr.185033.114 (2015). doi:10.1101/gr.185033.114

92. Gordân, R., Hartemink, A. J. & Bulyk, M. L. Distinguishing direct versus indirect transcription factor-DNA interactions. *Genome Res.* **19,** 2090–100 (2009).

93. de Vos, M. G. J., Dawid, A., Sunderlikova, V. & Tans, S. J. Breaking evolutionary constraint with a tradeoff ratchet. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 14906–11 (2015).

94. Baker, C. R., Tuch, B. B. & Johnson, A. D. Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 7493–7498 (2011).

95. Sayou, C. *et al.* A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* **343,** 645–8 (2014).

96. Maerkl, S. J. & Quake, S. R. Experimental determination of the evolvability of a transcription factor. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 18650–5 (2009).

97. Nadimpalli, S., Persikov, A. V. & Singh, M. Pervasive variation of transcription factor orthologs contributes to regulatory network evolution. *PLoS Genet.* **11,** e1005011 (2015).

Acknowledgments

J.A. and J.L.P. acknowledge support through the Forschungskredit program of the University of the Zurich, grants FK-14-076 and K-74301-04-01. J.L.P. acknowledges additional support through the Ambizione program of the Swiss National Science Foundation. A.W. acknowledges support through Swiss National Science Foundation grant 31003A_146137, as well as through the University Priority Research Program in Evolutionary Biology at the University of Zurich. We thank Sinisa Bratulic, Fahad Khalid, Andrés Moya, Yolanda Schaerli, and Macarena Toll-Riera for discussions and helpful comments on this manuscript.

Author Contributions

J.A., J.L.P., and A.W. designed research; J.A. and J.L.P. performed research; J.A., J.L.P., and A.W. analysed the data and wrote the paper.

Competing financial interests: The authors declare no competing financial interests.

Data availability: All data analysed during this study are available in public repositories; accession information is provided in Supplementary Tables 1 and 2. All data generated during this study are included in this published article (and its supplementary information files).

Figures

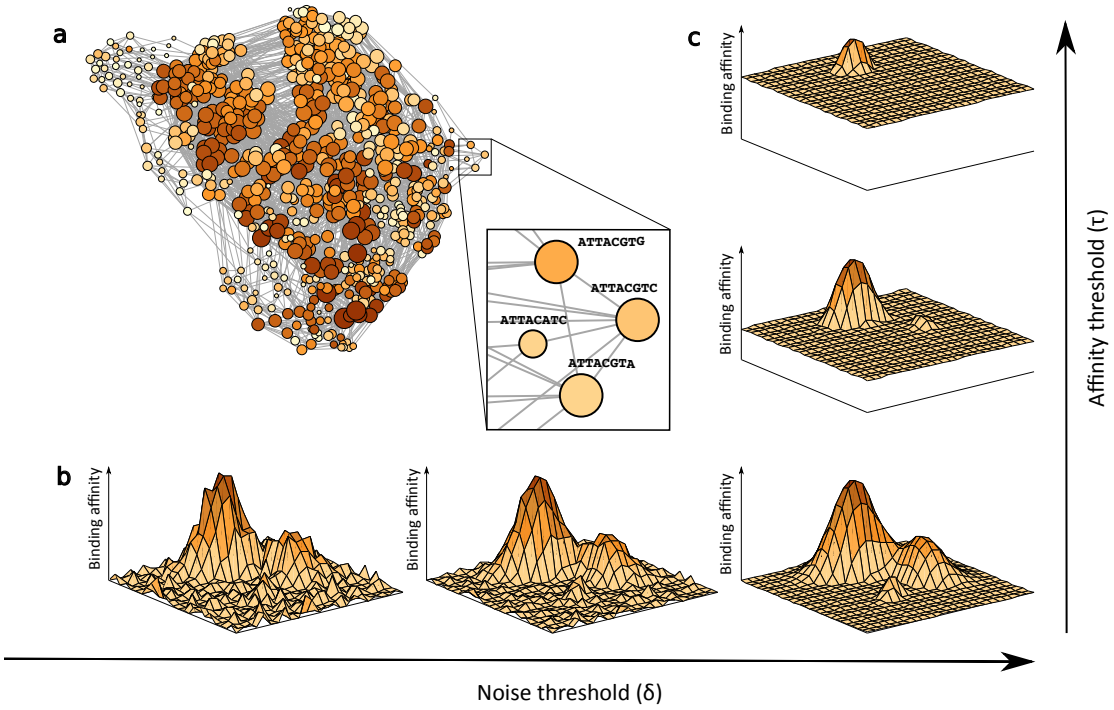


Figure 1. Adaptive landscapes of transcription factor binding affinity. (**A**) The largest connected component of the genotype network for the yeast TF Gcn4, visualized using a force-directed algorithm. Each vertex corresponds to a DNA sequence bound by Gcn4 (*E*-score > 0.35). The colour of a vertex indicates its binding affinity (darker = higher), i.e., the "elevation" of the landscape, whereas vertex size corresponds to the number of neighbouring sequences (bigger = more). The inset shows that two vertices are connected by an edge if their corresponding sequences are separated by a single small mutation (Materials and Methods). (**B**). To determine whether two sequences differ in their binding affinity, that is if they differ in their "elevation", we use a noise threshold δ, which accounts for experimental noise in the protein-binding microarray data. Increasing δ increases the "smoothness" of the landscape as shown here for a hypothetical landscape, where the *xy* plane represents genotype space, and the *z* axis represent binding affinity. (**C**) To delineate bound from unbound sequences we use a binding affinity threshold τ. As τ increases, the number of bound sequences decreases, thus reducing the size of the landscapes and pruning local peaks. In Supplementary sections 3.6.1 and 3.6.2, we assess the sensitivity of our main results to broadly varying δ and τ values.

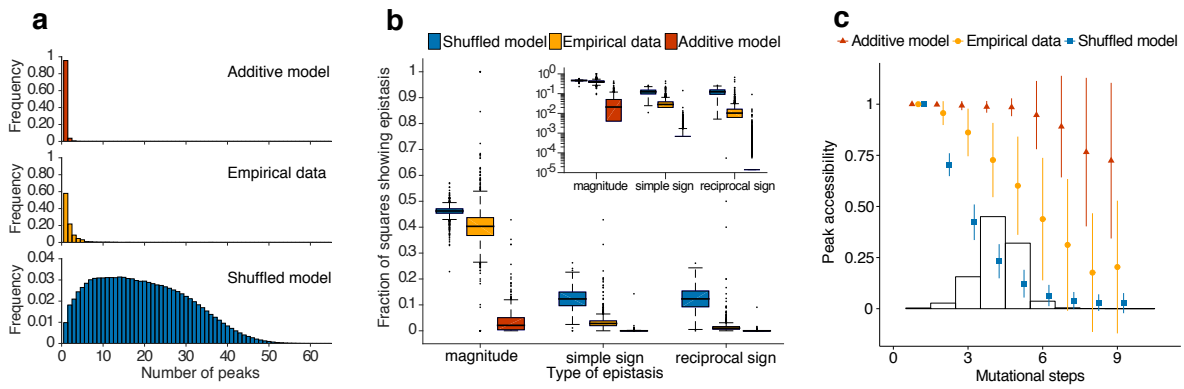


Figure 2. The navigability of adaptive landscapes of transcription factor binding affinity. (**A**) The distribution of the number of peaks for the 1,137 empirical adaptive landscapes (yellow) and the additive (red) and shuffled (blue) null models. (**B**) Boxplots of the fraction of squares showing magnitude, simple sign, and reciprocal sign epistasis for 1,137 adaptive landscapes. The thick horizontal line in the middle of each box represents the median of the data, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. For the shuffled model, the boxplot summarizes 1,137 data points, each of which is an average over 1000 shuffled landscapes. All pairwise differences are significant (Paired t-tests: $P < 2.2 \times 10^{-16}$). The inset shows the same data, but with a logarithmically scaled y-axis, which obscures the following numbers of data points with zero epistasis: magnitude epistasis (empirical: 9; additive: 184), simple sign epistasis (shuffled: 2; empirical: 53; additive: 681), and reciprocal sign epistasis (shuffled: 1; empirical: 98; additive: 853). The absence of the thick horizontal line indicates that the median of the distribution is below the lowest value of the logarithmically scaled y-axis. (**C**) For each of the 1,137 adaptive landscapes we show the mean (symbols) and standard deviation (error bars) of the fraction of accessible paths to the highest-affinity site in the landscape (i.e., peak accessibility). The histogram shows the distribution of mutational steps to the highest-affinity site for all sequences in all landscapes. For the shuffled model, each symbol represents the mean of 1,137 data points, each of which is an average over 1000 landscapes.

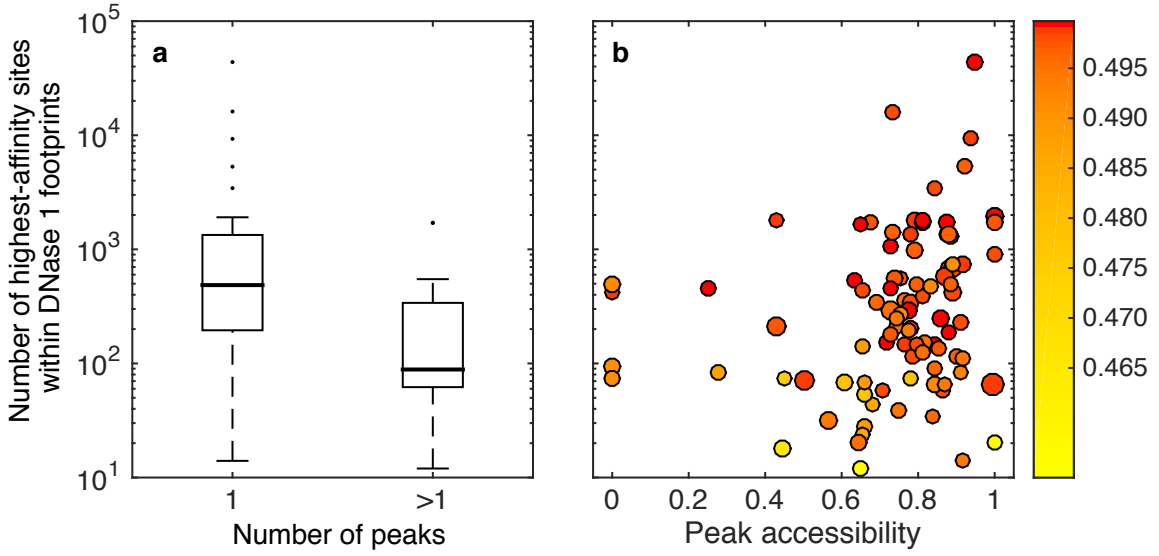


Figure 3. *In vivo* binding site abundance correlates with landscape navigability. The vertical axis of each panel indicates the abundance of a TF's highest-affinity site in protein-bound regions of the *M. musculus* genome in heart tissue (according to DNase I footprint data; Materials and Methods). In (**A**) the horizontal axis indicates the number of peaks, and the vertical axis classifies landscapes into two categories: single-peaked and multi-peaked. Global peak sequences from single-peaked landscapes are more abundant that those from multi-peaked landscapes (Wilcoxon rank-sum test, *P*-value = 1.42×10^{-6}). The thick horizontal line in the middle of each box represents the median of the data, while the bottom and top of each box represent the 25^{th} and 75^{th} percentiles, respectively. In (**B**) the horizontal axis shows peak accessibility through mutational paths of length 4 (Spearman's rank correlation coefficient = 0.27, *P*-value = 9.2×10^{-3}), which are the most abundant paths in our dataset (Fig. 2C). Each circle corresponds to a single TF expressed in heart tissue. Circle colour indicates the binding affinity of the TF's highest-affinity site (darker = higher; colour bar). Circle size corresponds to the TF's expression level (larger = higher, Materials and Methods). Note the logarithmic scale of the y-axis in both panels.

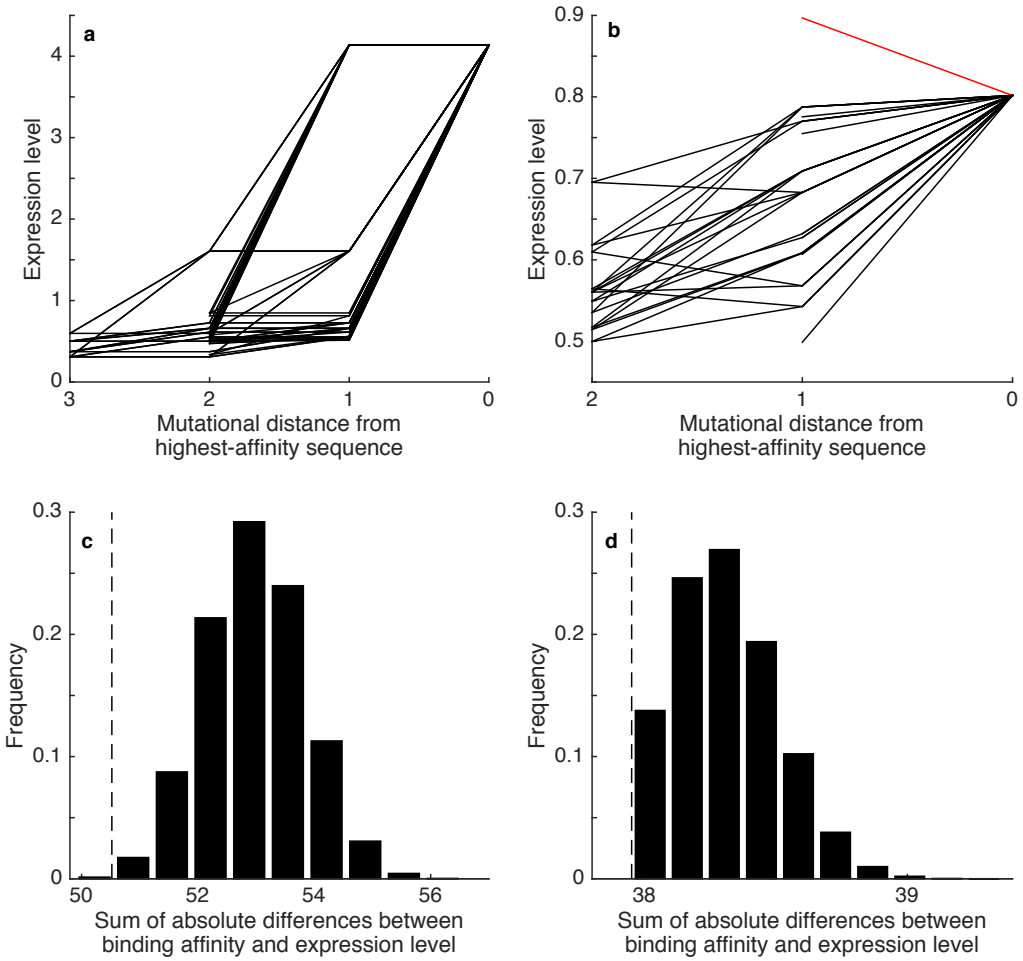


Figure 4. Gene expression increases along accessible mutational paths and reflects landscape topography. (**A, B**) Each line shows the expression levels of the sequences in an accessible mutational path to the highest-affinity sequence for the yeast TFs (**A**) Gcn4 and (**B**) Fhl1. Black lines denote accessible mutational paths in which gene expression increases monotonically. The red line denotes the single accessible mutational path in which gene expression does not increase monotonically. Note that while expression appears to decrease along some of the black lines in (**B**), it does not decrease beyond the noise threshold δ necessary to classify the trend as a true expression decrease (δ = 0.7 for Gcn4 and δ = 0.1 for Fhl1; Material and Methods). Note also that the highest-affinity sequence does not necessarily correspond to the sequence with the highest level of expression (red line in (**B**)). (**C, D**) Vertical dashed lines indicate the observed sum of the absolute differences between binding affinity and gene expression, and the black bars show the null distribution of this sum for 10^5 random permutations of the gene expression values for the TFs (**C**) Gcn4 and (**D**) Fhl1.

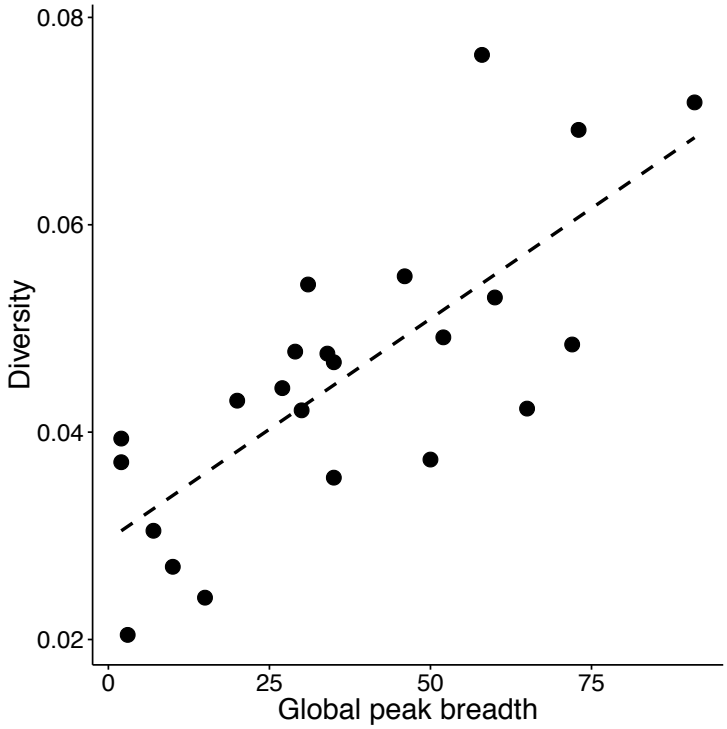
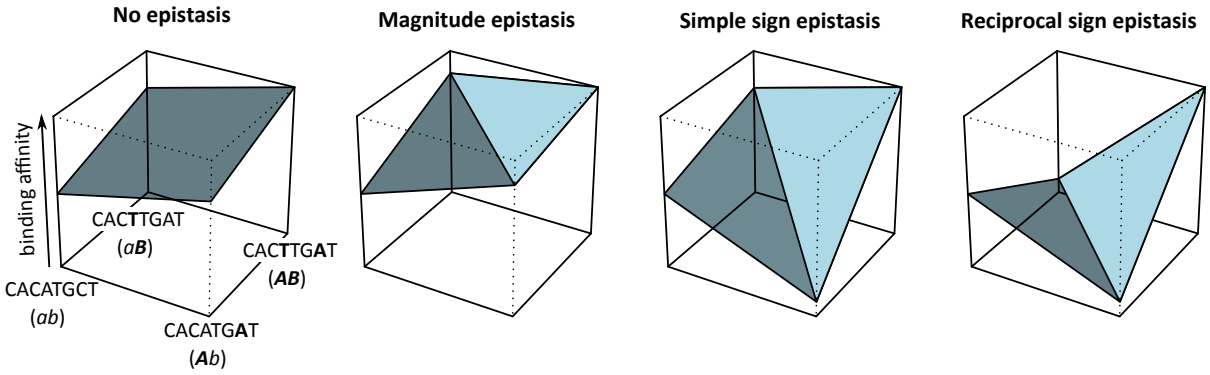


Figure 5. **Global peak breadth influences the diversity of TF binding sites in the yeast genome.** Scatter-plot showing the relationship for 23 TFs between the number of binding sites in the global peak (i.e., global peak breadth) of a particular TF and the average diversity over all its polymorphic binding sites in nineteen strains of *S. cerevisae* (Materials and Methods; Spearman's rank correlation coefficient = 0.73, *P*-value = 8.9×10^{-5}). The dashed line represents the best linear-regression fit to the data.



Box 1. Epistasis. A square in a genotype network connects a "wild type" sequence (*ab*) to a double mutant (***AB***) through two single mutants (***A****b* and *a****B***). The left-most panel shows an example with binding sites of length eight where there is no epistasis: The binding affinity of the double mutant is simply the addition of the affinity contributions of the two single mutants. That is, the mutation *a* to ***A*** has the same effect on binding affinity in the two different genetic backgrounds (*b* and ***B***). *Magnitude epistasis* occurs when the magnitude (but not the sign) of a mutation's effect on binding affinity depends on the genetic background. *Simple sign epistasis* occurs when one single mutant has a lower binding affinity than both the wild type and the double mutant, while the other single mutant has an affinity that is intermediate to the wild type and double mutant[64]. *Reciprocal sign epistasis* occurs when both mutations decrease affinity independently, but increase affinity in combination[29]. To account for the noise in our data, we only consider a single mutant's binding affinity to be lower than that of the wild type if the affinity difference exceeds a threshold value δ, which is derived from the empirical data and is specific to each TF (Materials and Methods). We refer generically to both simple sign epistasis and reciprocal sign epistasis as *sign epistasis*. For all squares, we quantified epistasis along a single axis of the square by designating the highest-affinity sequence as the double mutant (Materials and Methods). The figure in this box has been adapted from Poelwijk *et al.*[58].

Supplementary Information for

"A thousand empirical adaptive landscapes and their navigability"

José Aguilar-Rodríguez†, Joshua L. Payne†, and Andreas Wagner*

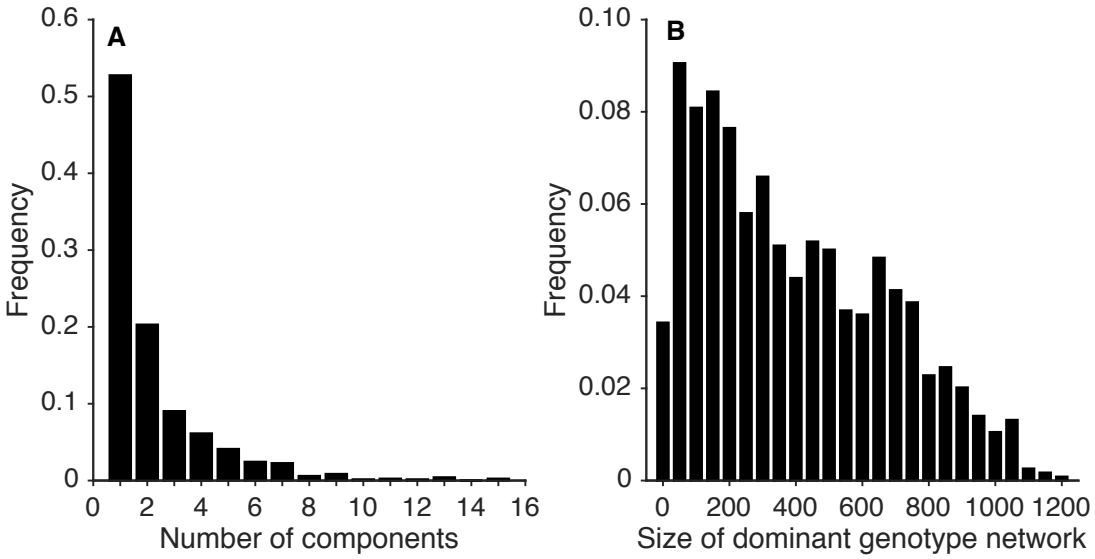
†These authors contributed equally to this work

*Corresponding author: andreas.wagner@ieu.uzh.ch

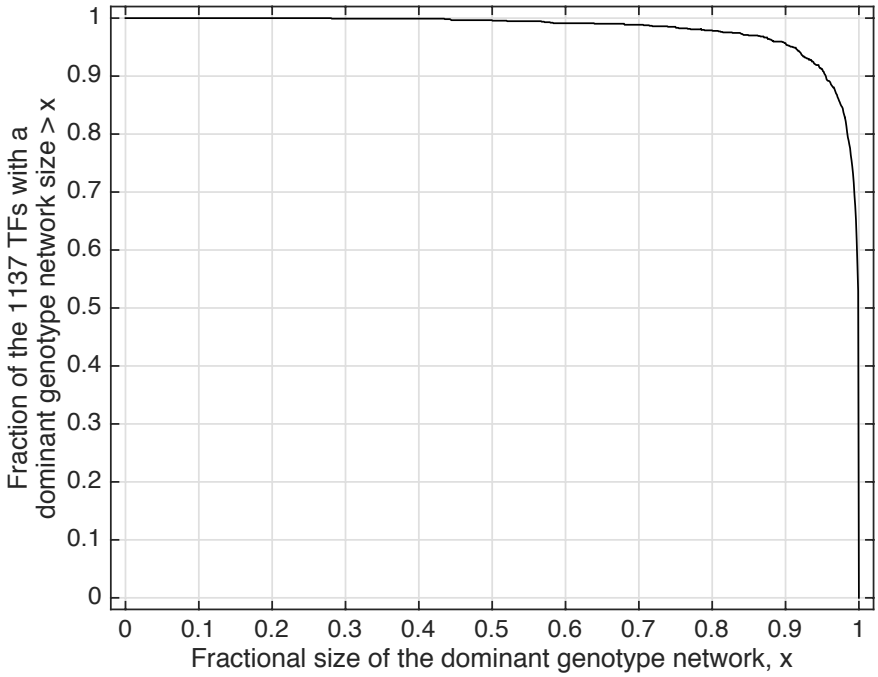
Contents

1. Supplementary figures **33**
2. Supplementary tables **66**
3. Supplementary results **67**
3.1. Summary statistics of genotype networks **67**
3.2. Global peaks are usually organized into broad plateaus **67**
3.3. Why epistasis occasionally appears in the additive null model **68**
3.4. Sign epistasis preferentially occurs among nucleotides that are near one another in the binding site **68**
3.5. Peak accessibility decreases when unbound sequences are included **69**
3.6. Sensitivity analyses **70**
3.6.1. Our observations are insensitive to broadly varying thresholds for noise filtering 70
3.6.2. Our observations are insensitive to broadly varying affinity thresholds for delineating bound from unbound sequences 71
3.6.3. Our observations are consistent across DNA binding domains 73
3.6.4. Our observations are consistent across TFs that bind shorter or longer sequences than eight nucleotides 74
3.6.5. Peak breadth is sensitive to the use of E-scores as a quantitative phenotype 75
3.7. The *in vivo* relationship between landscape navigability and the abundance of binding sites is not driven by binding affinity or by information content **75**
3.8. Our measures of epistasis for bound sequences are conservative **76**
4. Supplementary discussion **77**
4.1. Caveats **79**
5. Supplementary references **81**

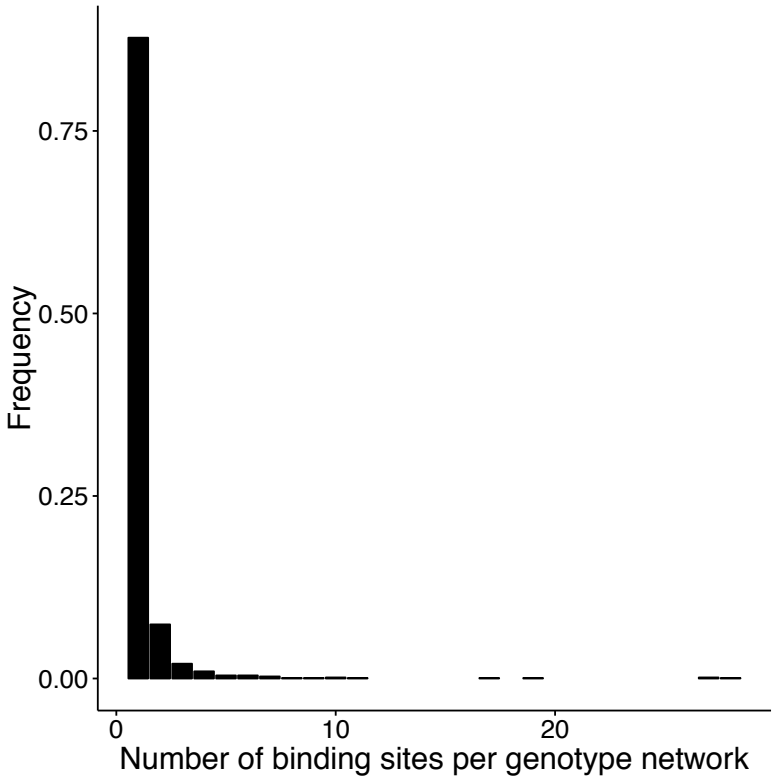
1. Supplementary figures



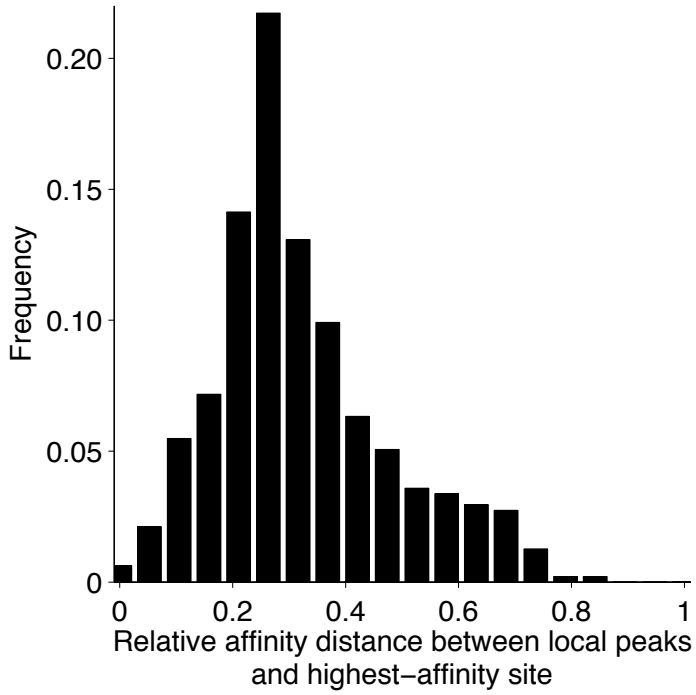
Supplementary Figure 1. Structural properties of the genotype networks of 1,137 TFs. The distribution of (**A**) the number of components and (**B**) the size of the largest (i.e., dominant) connected component in the genotype networks of 1,137 TFs.



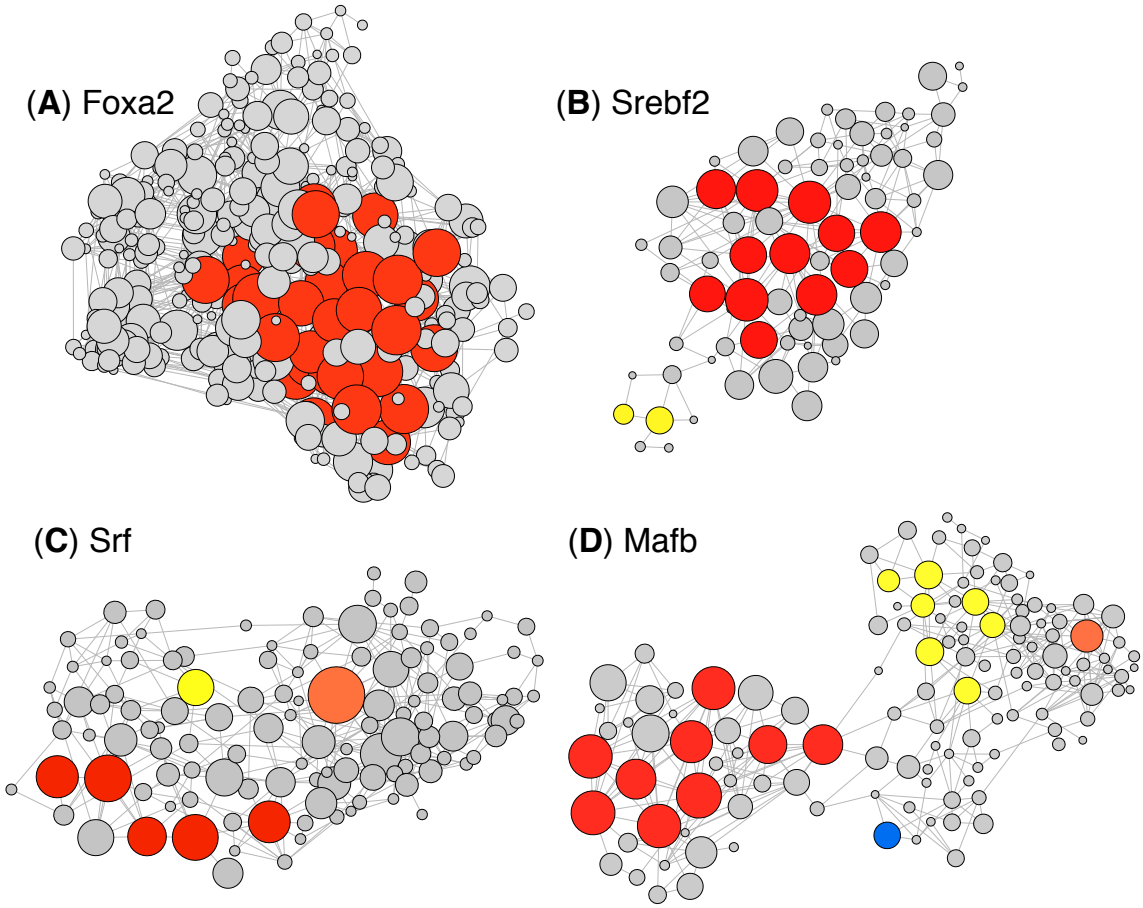
Supplementary Figure 2. Cumulative distribution of the fraction of TFs in our dataset that have a dominant genotype network comprising at least *x*% of the bound sequences. This figure shows that for the vast majority of TFs, the vast majority of bound sequences are in the dominant genotype network. For example, 96% of the 1137 dominant genotype networks comprise at least 90% of the sequences that bind the TF.



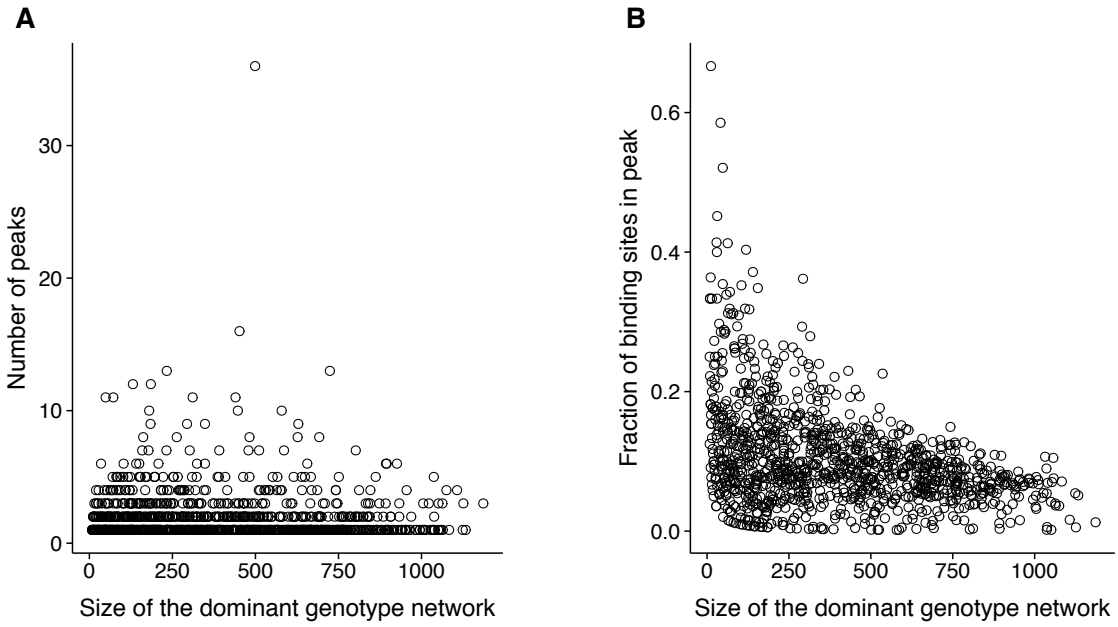
Supplementary Figure 3. The number of sequences per non-dominant genotype network is very small. Black bars show the distribution of the number of sequences contained in the non-dominant genotype networks of 1,137 TFs. 87.8% of these networks comprise a single sequence.



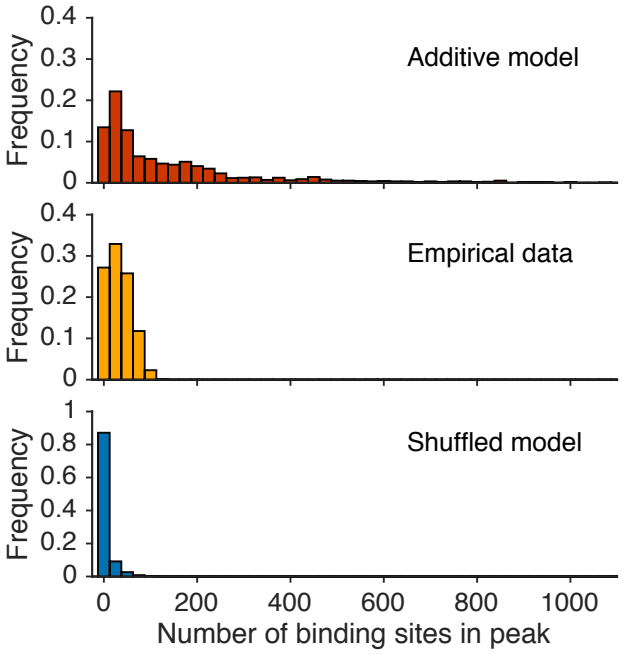
Supplementary Figure 4. Multiple peaks of unequal height. The distribution of the average normalized difference in binding affinity between the highest-affinity sequence in a landscape and the mean binding affinity of the landscape's local peaks, for 478 multi-peaked landscapes. This distance D_i of a local peak i is calculated as $E_g - \frac{1}{N_i}\sum_j E_{ij}$, where E_g is the binding affinity of the highest-affinity site in the global peak, N_i is the number of sites in local peak i, and E_{ij} is the binding affinity of site j in local peak i. All values are normalized by 0.15, which is the maximum range of affinity when the binding threshold is $\tau = 0.35$. Non-normalized values range from 1.3×10^{-3} to 0.125.



Supplementary Figure 5. Visualization of adaptive landscapes that have different numbers of peaks. Representative examples of landscapes with (**A**) one, (**B**) two, (**C**) three, and (**D**) four peaks. Vertices represent bound sequences, vertex sizes represent binding affinity, and vertex colours represents peak membership. Red indicates a global peak and grey indicates bound sequences that do not belong to any peak. Notice that peaks are usually organized into broad plateaus comprising binding sites with similar binding affinity (Supplementary Fig. 7), that is with values in affinity above a threshold E_g - δ, where E_g is the affinity of the highest-affinity site in the peak and δ is the noise threshold (Material and Methods).



Supplementary Figure 6. Small landscapes are not necessarily more navigable than large landscapes. The relationship between the size of the dominant genotype network and (**A**) the number of peaks (Spearman's rank correlation coefficient = -0.07, *P*-value < 1.9×10^{-2}), and (**B**) the relative size of the global peak for 1,137 TFs (Spearman's rank correlation coefficient = -0.34, *P*-value < 2.2×10^{-16}).



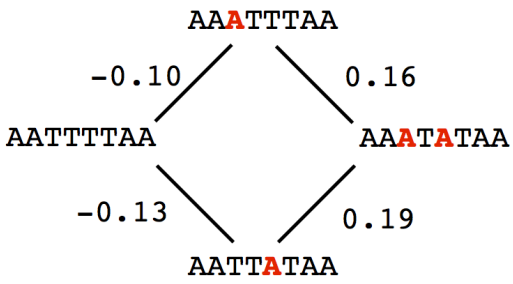
Supplementary Figure 7. Global peaks are usually organized into broad plateaus. The distribution of the number of binding sites per global peak for the 1,137 empirical adaptive landscapes (yellow) and the additive (red) and rugged (blue) null models.

(A)

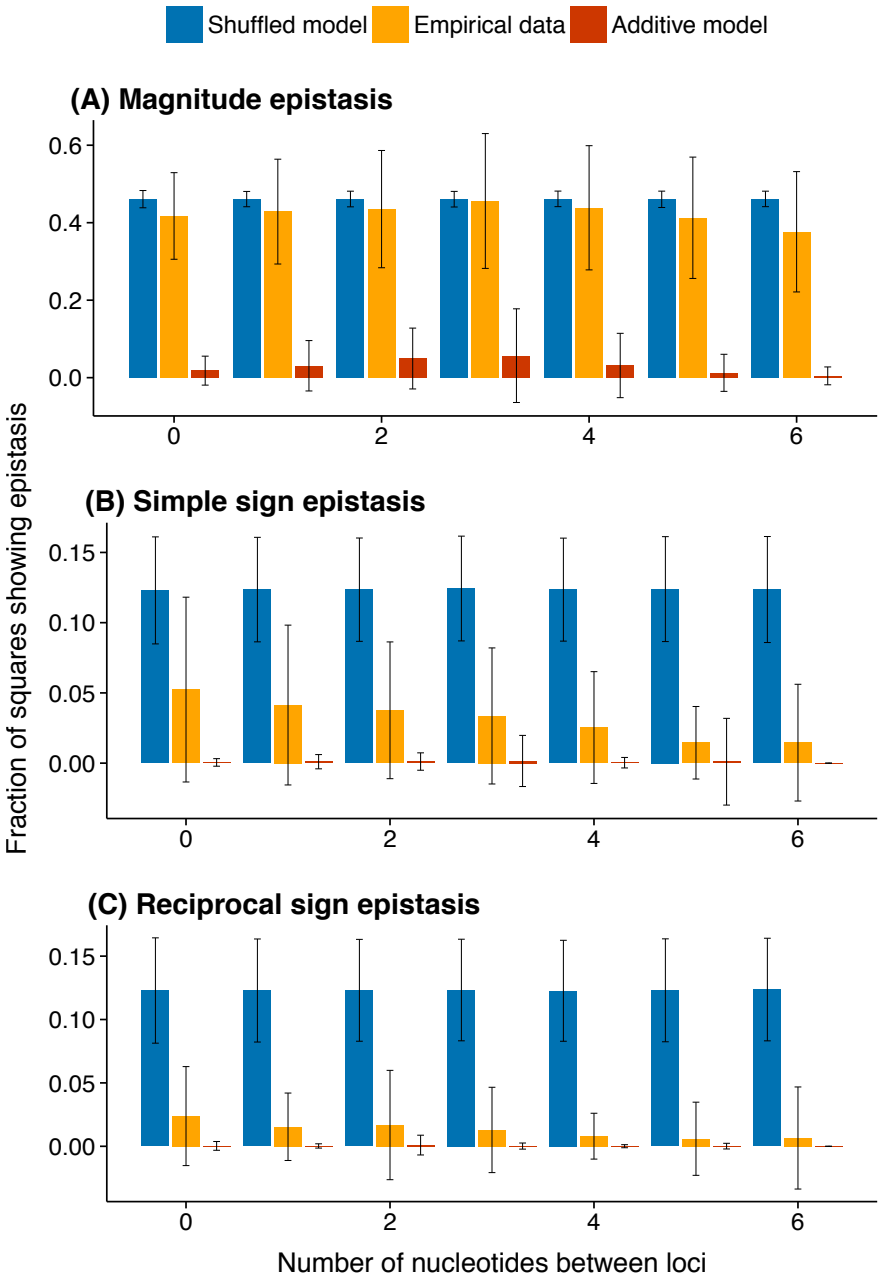
(B)

A:	0.177	0.160	0.852	0.961	0.056	0.930	0.019	0.041	0.233	0.339	0.285	0.572	0.285	0.342
C:	0.428	0.324	0.026	0.003	0.004	0.005	0.014	0.087	0.027	0.352	0.134	0.078	0.184	0.230
G:	0.229	0.032	0.073	0.010	0.007	0.007	0.006	0.039	0.519	0.184	0.268	0.190	0.262	0.153
T:	0.165	0.482	0.047	0.024	0.931	0.057	0.959	0.832	0.218	0.123	0.311	0.159	0.268	0.272
I(i):	0.118	0.384	**1.183**	**1.724**	**1.589**	**1.579**	**1.703**	**1.101**	0.399	0.120	0.062	0.375	0.019	0.058
Pos:	1	2	**3**	**4**	**5**	**6**	**7**	**8**	9	10	11	12	13	14

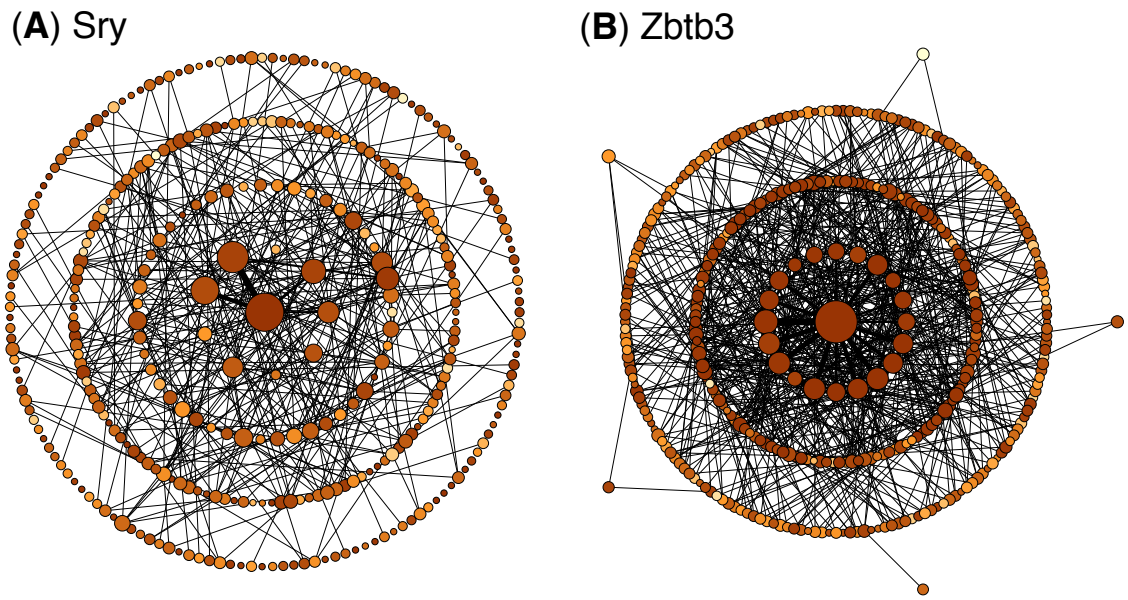
(C)



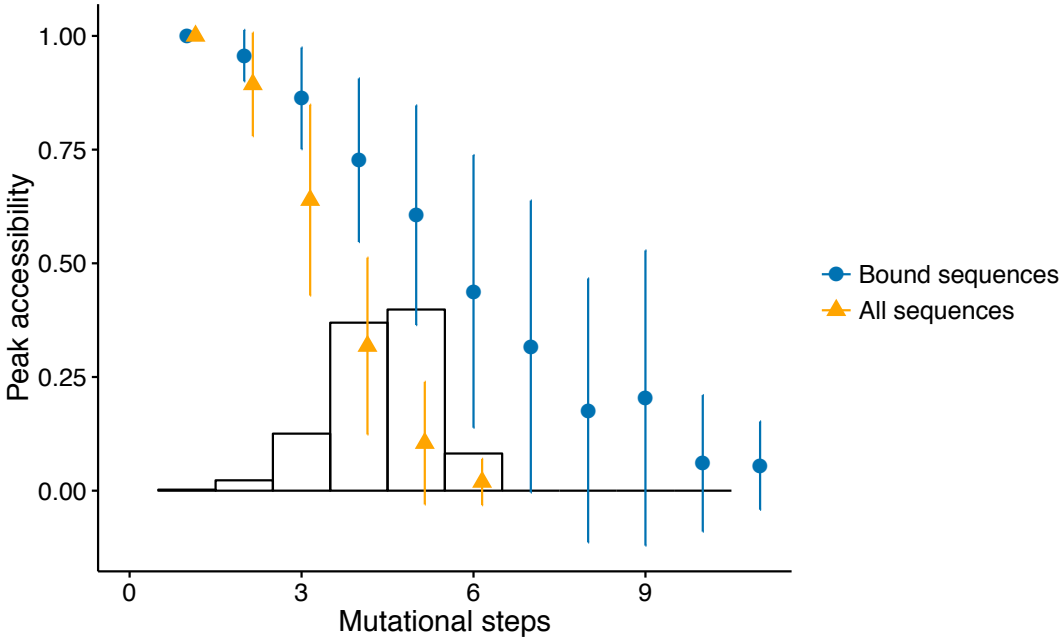
Supplementary Figure 8. Why epistasis occasionally appears in the additive null model. (**A**) Sequence logo of (**B**) the position weight matrix (PWM) for the mouse TF Arid5a downloaded from UniPROBE[20]. The information content (Eq. 8, Materials and Methods) per position is also shown. Positions in bold face indicate the largest contiguous subset of positions with information content above 0.5, which we use to calculate the effective width of a PWM (Supplementary section 3.6.4). (**C**) A square that exhibits reciprocal sign epistasis due to the sliding-window approach for scoring sequences in the additive null model. Edge labels indicate the change in binding affinity between a pair of sequences. For the sequence AATTTTAA, the highest-scoring match occurs when it is aligned with positions 3 through 10 in the PWM, whereas for the other three sequences, the highest-scoring match occurs when they are aligned with positions 2 through 9 in the PWM. In this example, the noise threshold δ is 0.09 and the magnitude of epistasis ε (Eq. 3, Materials and Methods) is 0.29.



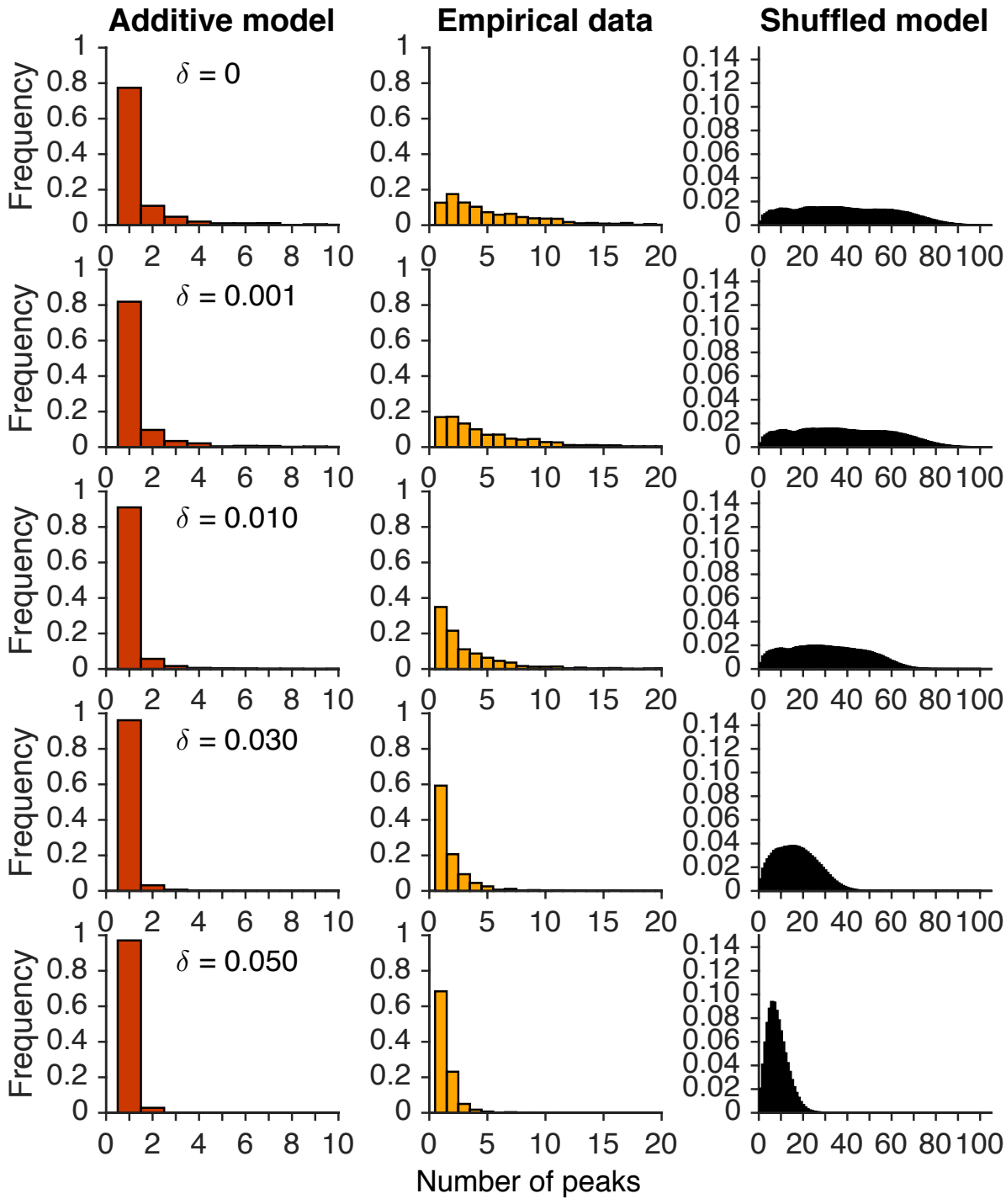
Supplementary Figure 9. Sign epistasis preferentially occurs among nucleotides that are near one another in the binding site. Bar plots of the levels of (**A**) magnitude (**B**) simple sign, and (**C**) reciprocal sign epistasis as a function of the distance between the two mutations in the TF binding site. The height of the bars represents the median of the data, while the error bars represent the standard deviation. For each TF, the level of epistasis plotted for the random landscapes is an average of 1,000 shuffled landscapes. To facilitate a direct comparison with Jolma *et al.*[30], squares that include indels are excluded from this analysis.



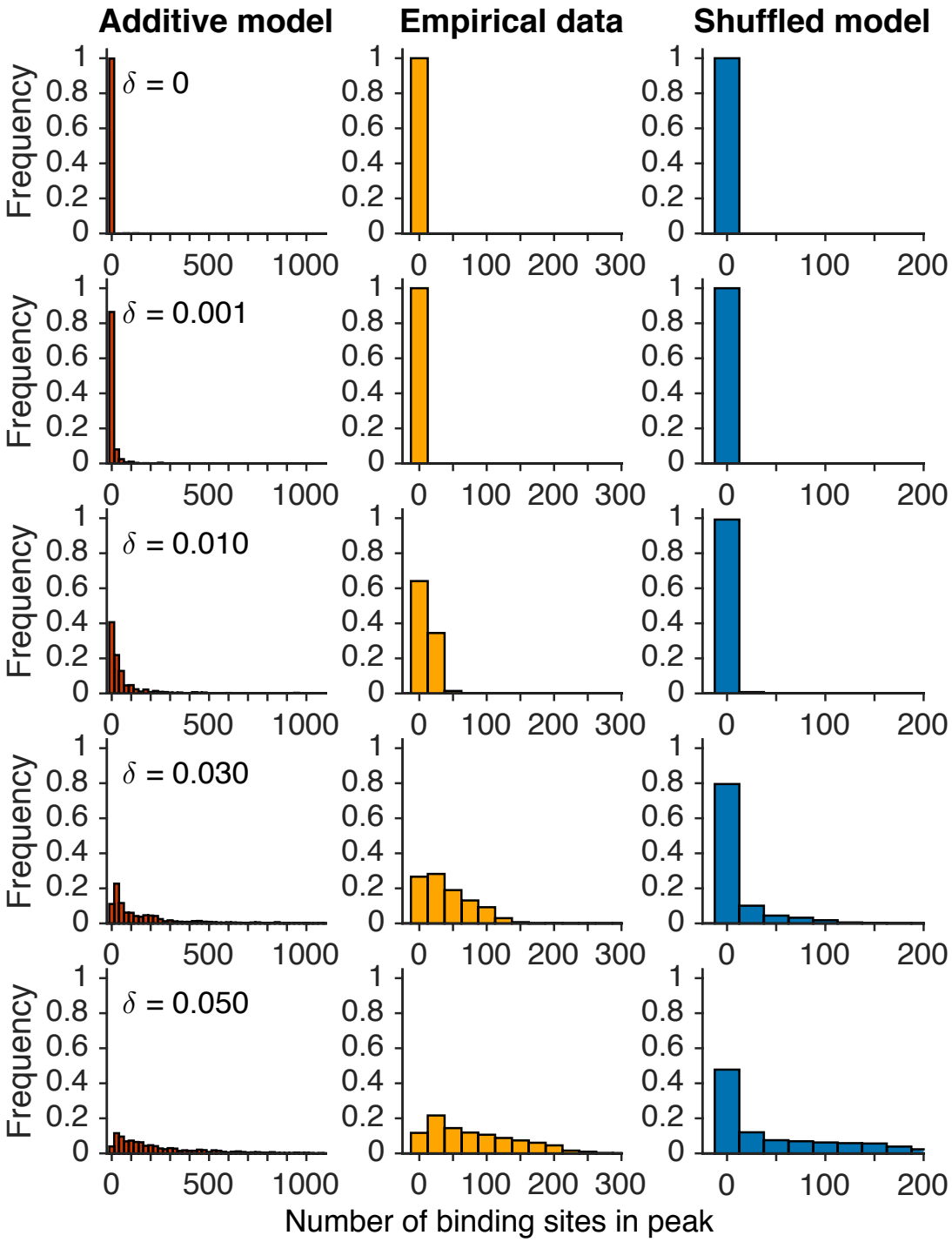
Supplementary Figure 10. Visualization of two global peaks that vary in their accessibility. The murine TFs (**A**) Sry and (**B**) Zbtb3 bind a similar number of sequences (356 and 352, respectively), but these sequences differ dramatically in their ability to evolve higher binding affinity through small mutations. Each visualization depicts the genotype networks of all bound sequences that are within four mutations of the highest-affinity site (centre vertex), i.e., the summit of the landscape, in a layout where each concentric ring corresponds to a distinct mutational distance from the summit. Edge width and vertex size indicate the number of accessible mutational paths (i.e., paths that access the summit via monotonic increases in binding affinity) that include the edge or vertex. Vertex color indicates binding affinity, as in Fig. 1A.



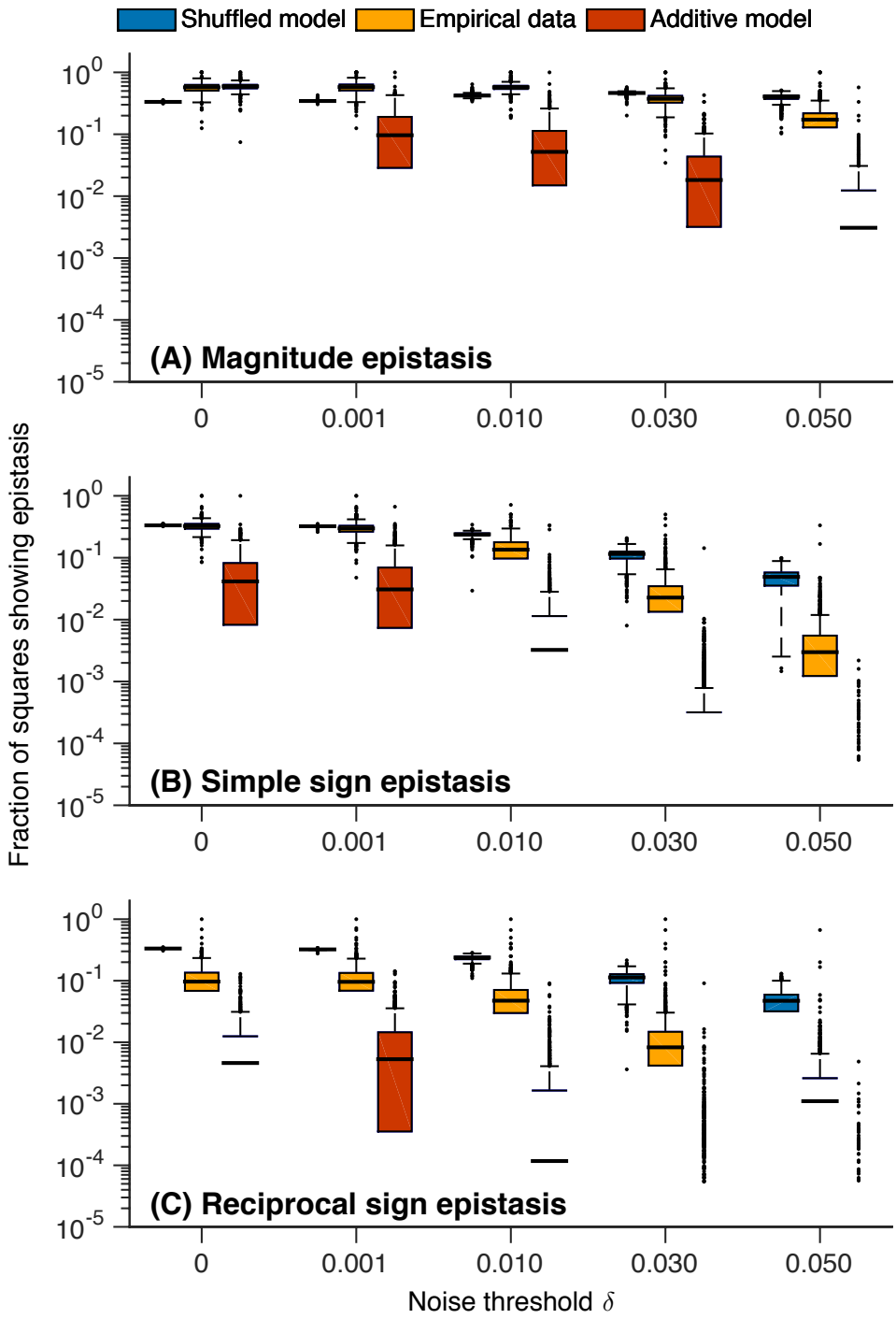
Supplementary Figure 11. Peak accessibility is reduced by the inclusion of unbound sequences. Data correspond to all 1,137 TFs. Symbols show the mean, and error bars show the standard deviation. The histogram shows the distribution of the lengths of the shortest mutational paths from bound sequences (i.e., $\tau > 0.35$) to the global peaks of the 1,137 TFs' landscapes, when unbound sequences (i.e., $\tau \leq 0.35$) are considered. Note that there is no shortest mutational path longer than 6 mutational steps in this case, in contrast to when unbound sequences are not considered (cf. Figure 2C).



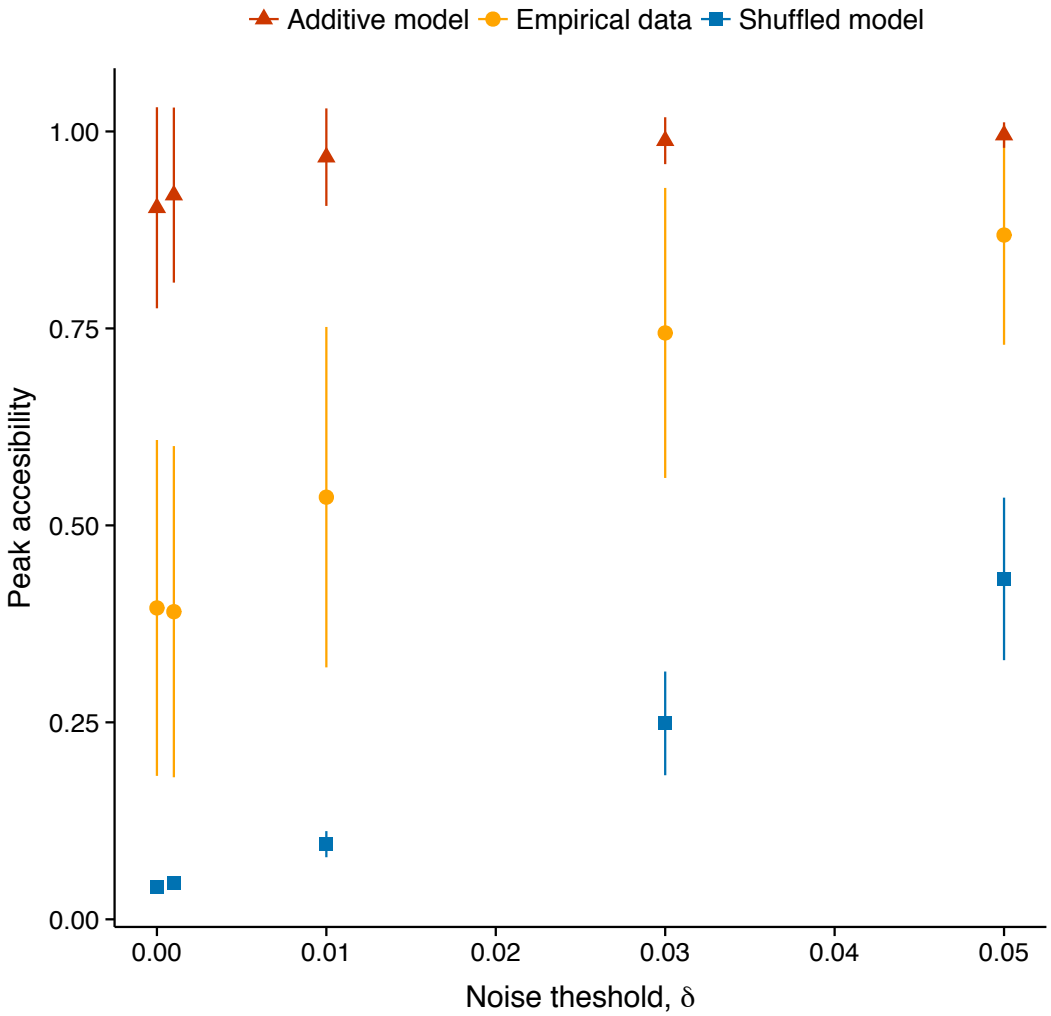
Supplementary Figure 12. The number of peaks per landscape decreases as the noise threshold increases. These data represent a sensitivity analysis of the trends presented in Fig. 2A. Each panel shows the distribution of the number of peaks per landscape for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to a different noise threshold δ.



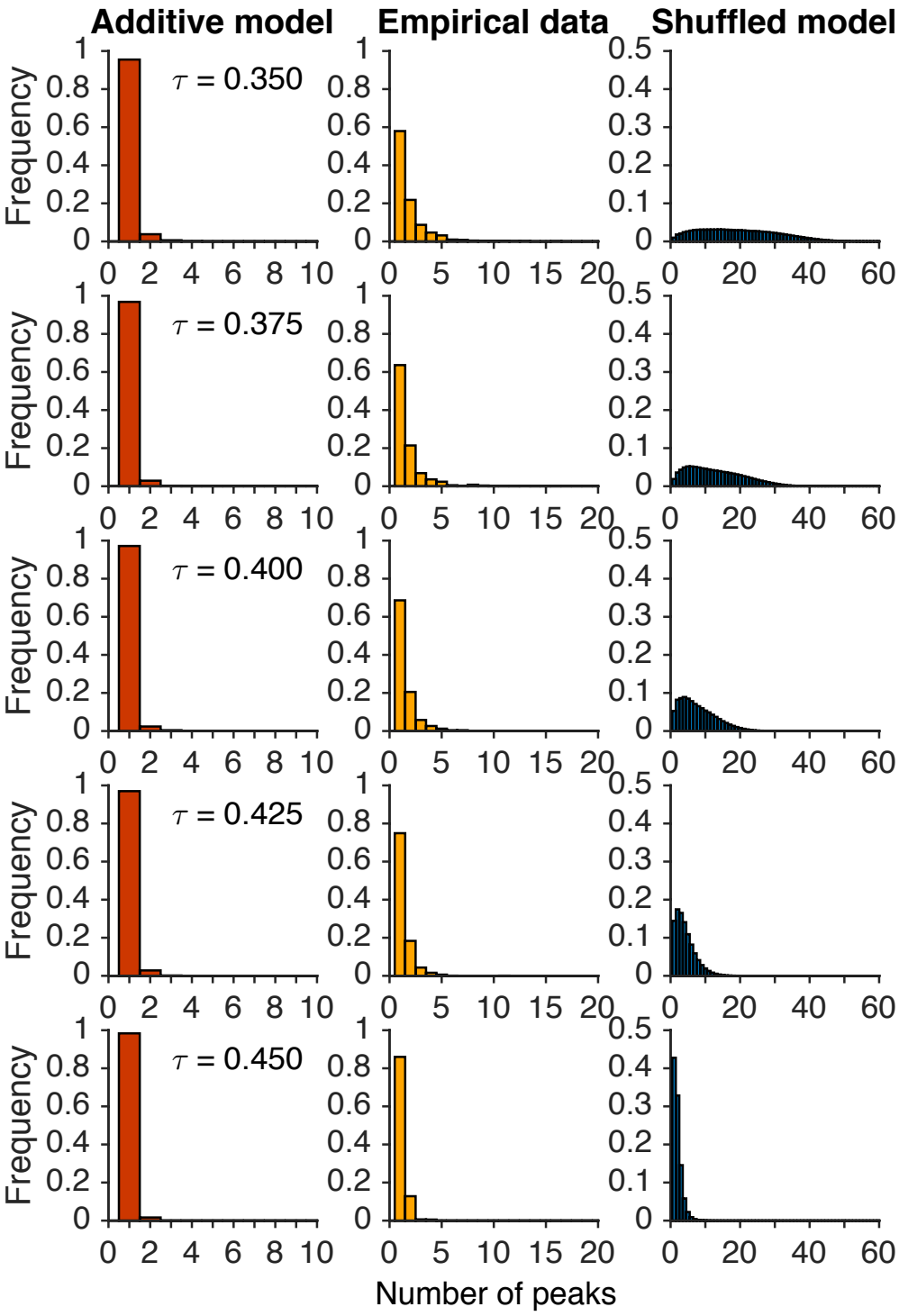
Supplementary Figure 13. Global peak breadth increases as the noise threshold increases. These data represent a sensitivity analysis of the trends presented in Supplementary Fig. 7. Each panel shows the distribution of the number of binding sites per global peak (i.e., global peak breadth) for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to a different noise threshold δ.



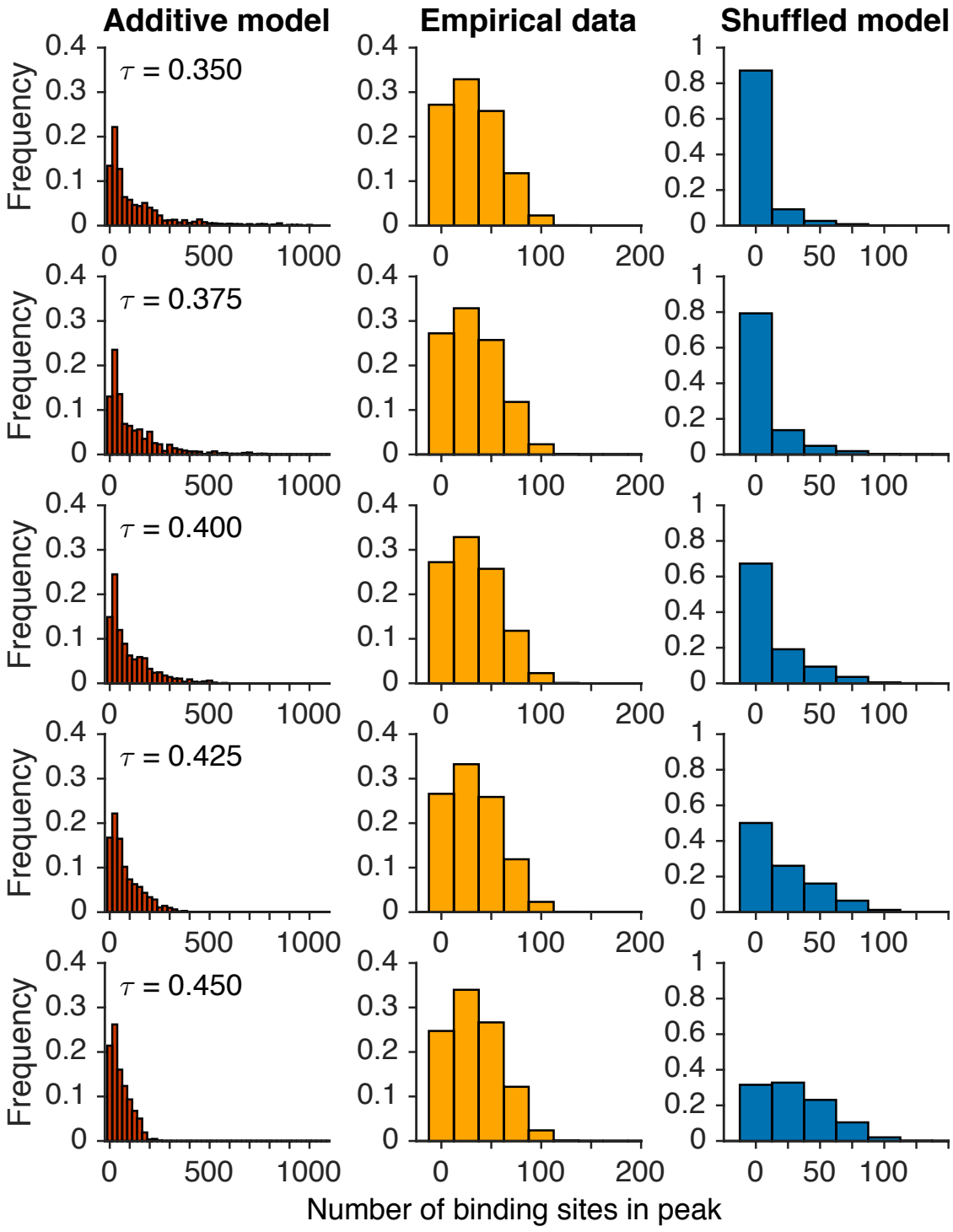
Supplementary Figure 14. Sign epistasis in the empirical data is intermediate to that of the additive and shuffled models for all noise thresholds. These data represent a sensitivity analysis of the trends presented in Fig. 2B. Boxplots of (**A**) magnitude epistasis, (**B**) simple sign epistasis, and (**C**) reciprocal sign epistasis as a function of the binding affinity threshold τ. The thick horizontal line in the middle of each box represents the median of the data, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. Note the logarithmic scale of the y-axes, which obscures data points with zero epistasis. The absence of the thick horizontal line indicates that the median of the distribution is below the lowest value of the y-axis.



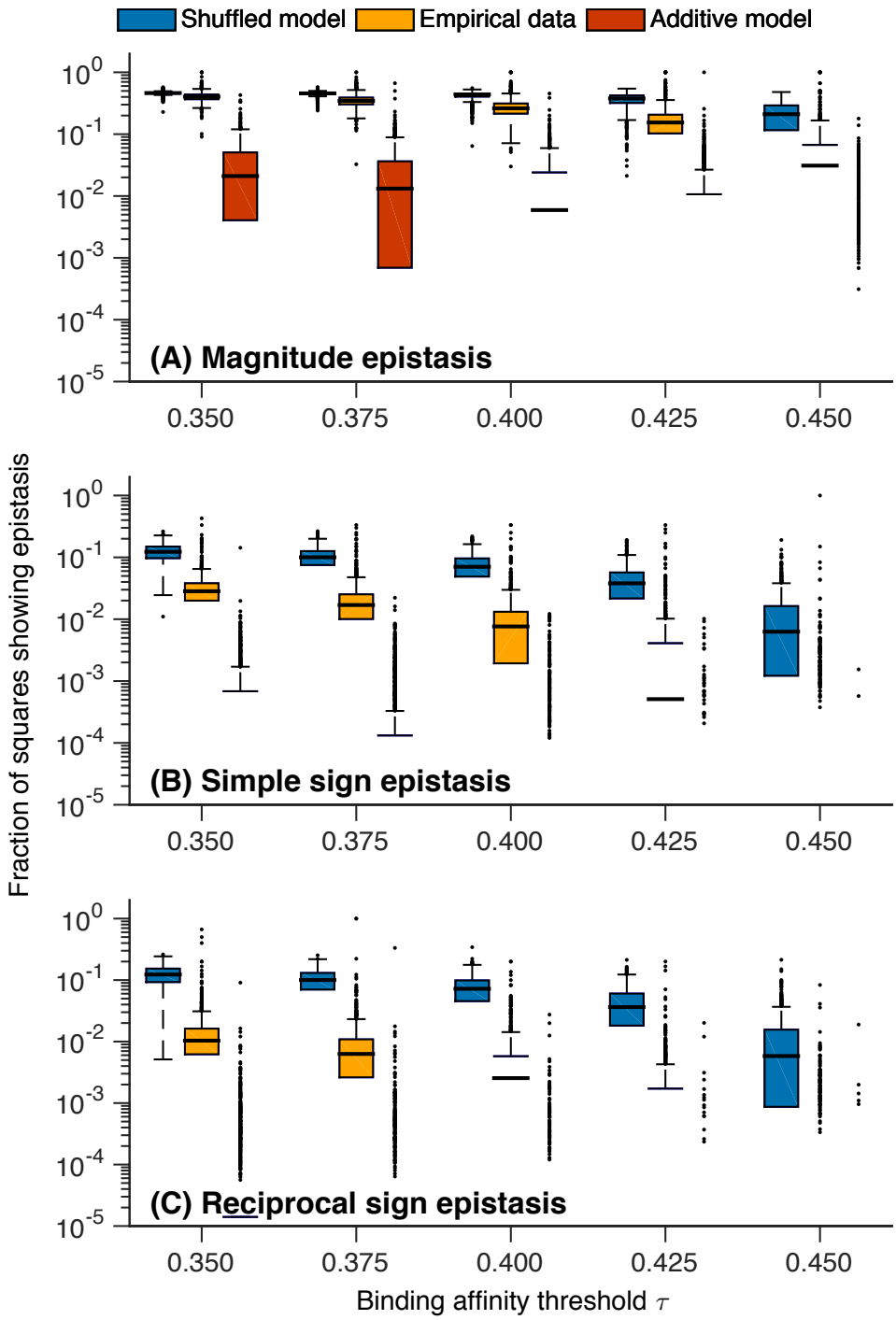
Supplementary Figure 15. Accessibility of the highest affinity site increases as the noise threshold increases. These data represent a sensitivity analysis of the trends presented in Fig. 2C for mutational paths of length 4, which are the most abundant in our dataset. Data points depict the mean peak accessibility in the empirical data and the two null models (see legend) as a function of the noise threshold δ. Error bars depict one standard deviation.



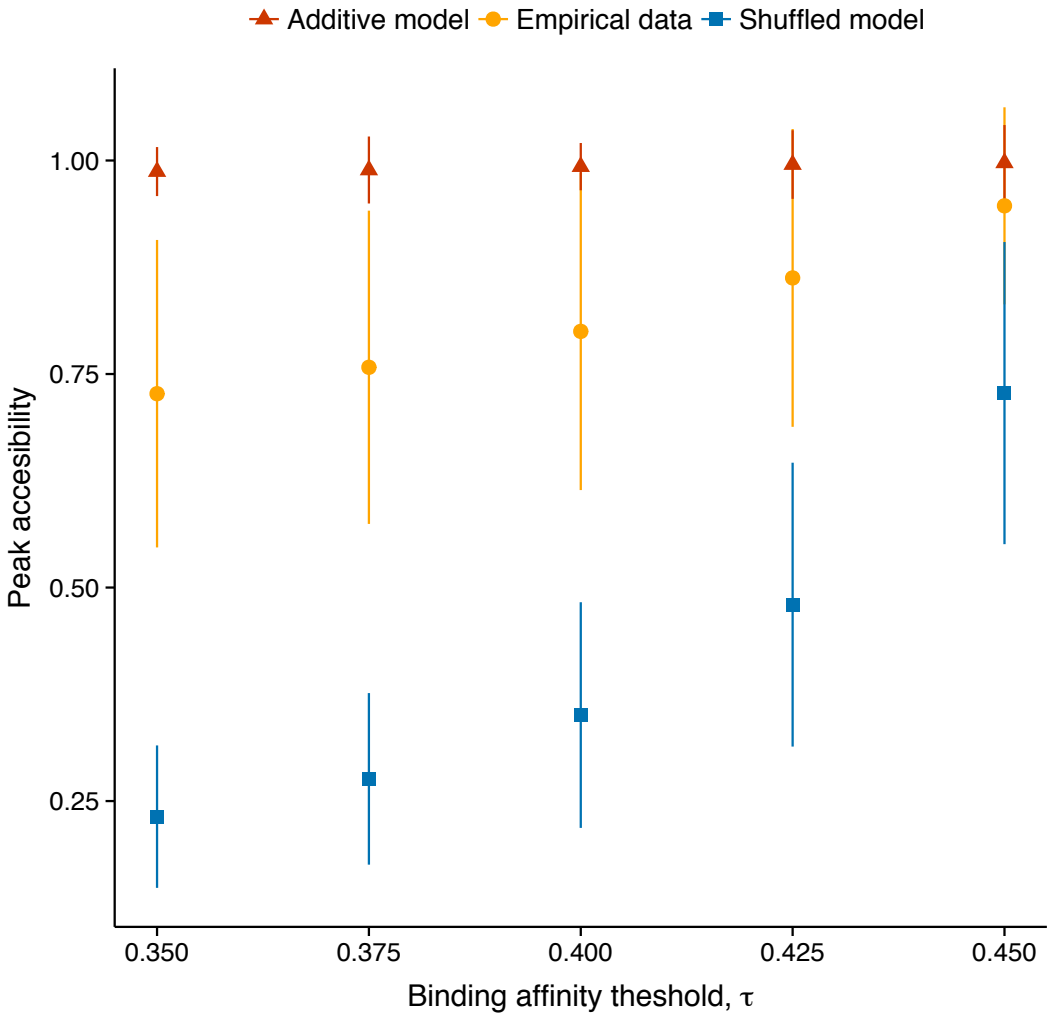
Supplementary Figure 16. The number of peaks per landscape decreases as the binding affinity threshold increases. These data represent a sensitivity analysis of the trends presented in Fig. 2A. Each panel shows the distribution of the number of peaks per landscape for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to a different binding affinity threshold τ.



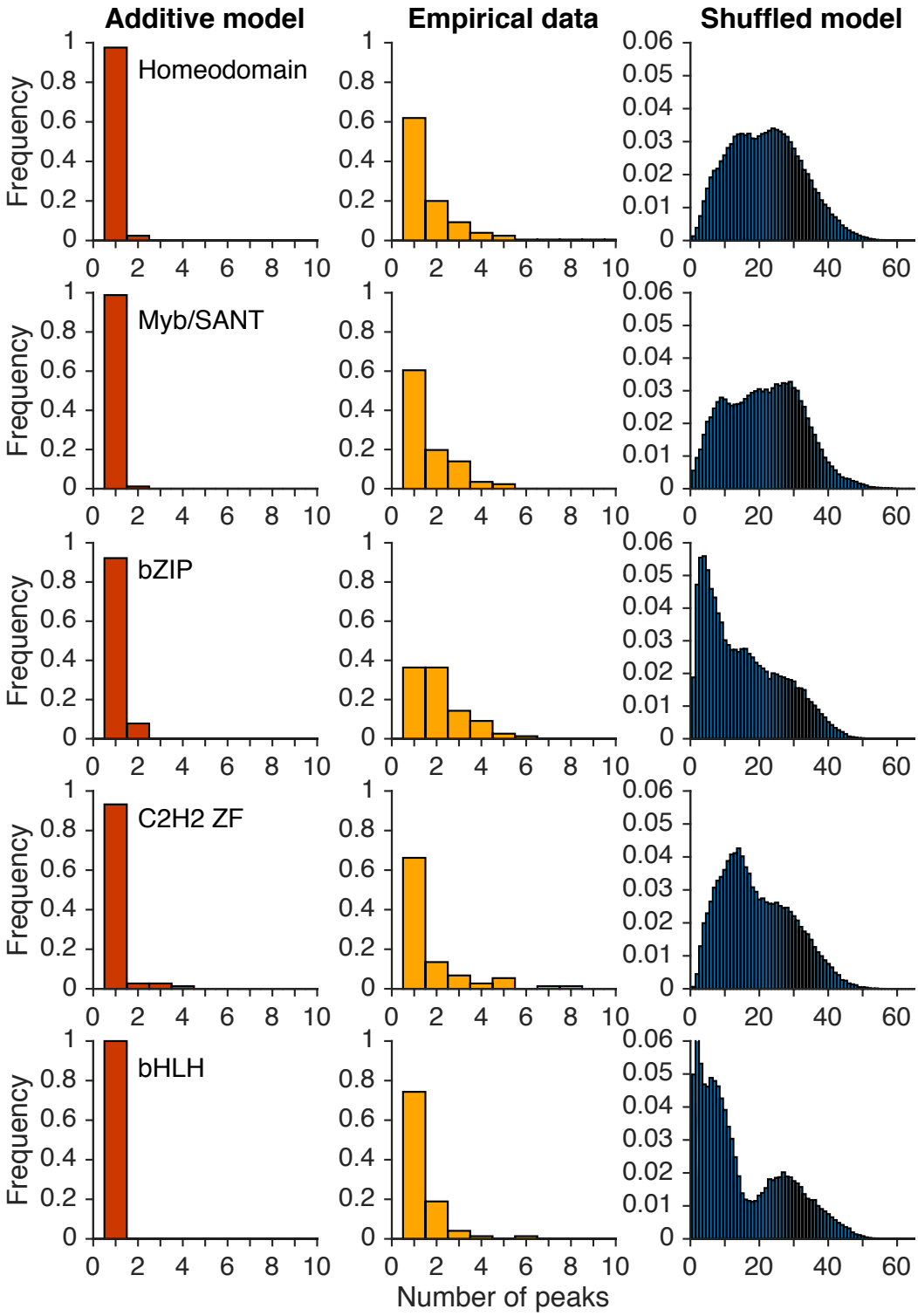
Supplementary Figure 17. The empirical distribution of global peak breadth does not vary with the binding affinity threshold. These data represent a sensitivity analysis of the trends presented in Supplementary Fig. 7. Each panel shows the distribution of the number of binding sites per global peak (i.e., global peak breadth) for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to a different binding affinity threshold τ.



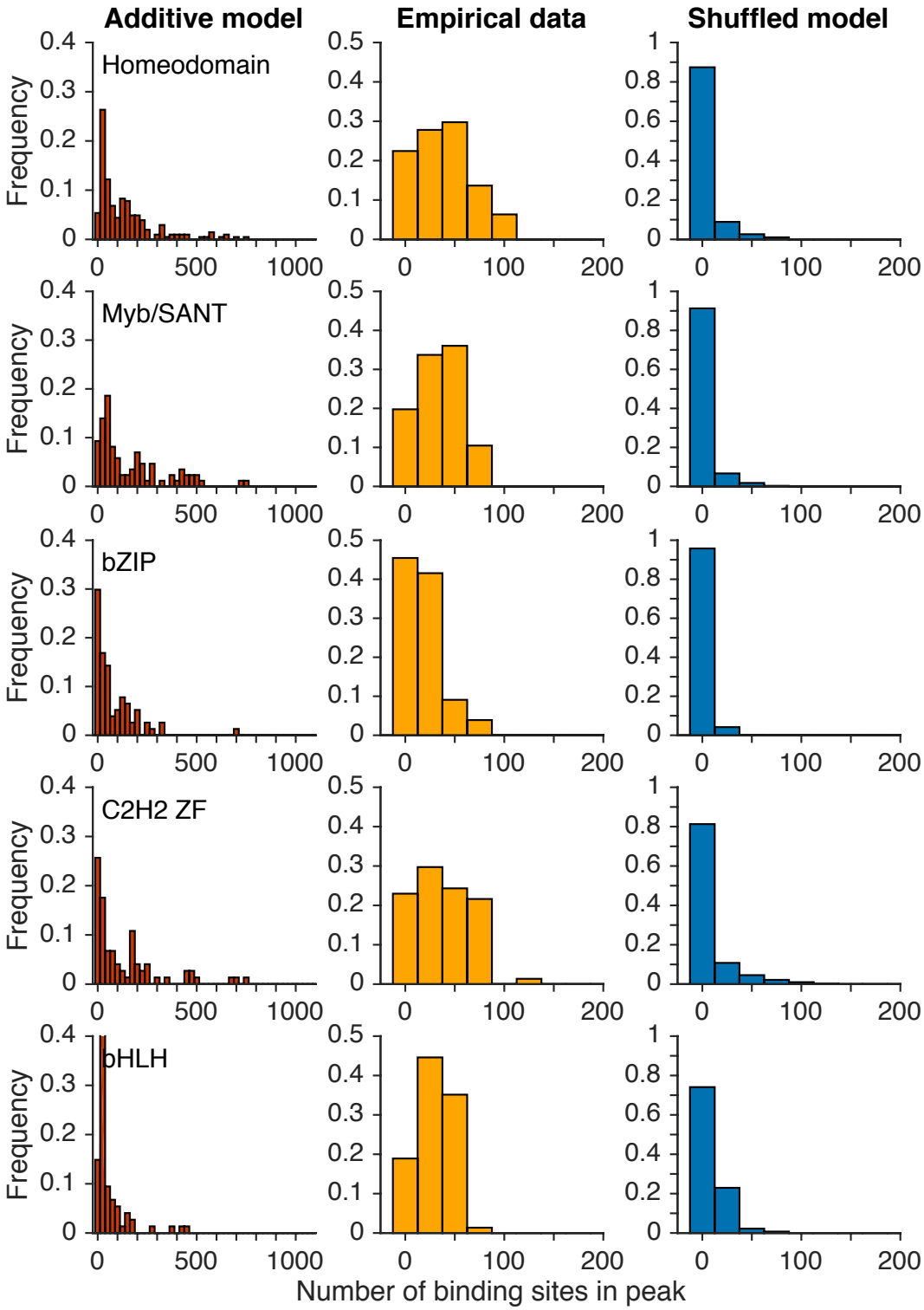
Supplementary Figure 18. Epistasis in the empirical data is intermediate to that of the additive and shuffled models for all affinity thresholds. These data represent a sensitivity analysis of the trends presented in Fig. 2B. Boxplots of (**A**) magnitude epistasis, (**B**) simple sign epistasis, and (**C**) reciprocal sign epistasis as a function of the binding affinity threshold τ. The thick line in the middle of each box represents the median of the data, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. Note the logarithmic scale of the y-axes, which obscures data points with zero epistasis. The absence of the thick line indicates that the median of the distribution is below the lowest value of the y-axis.



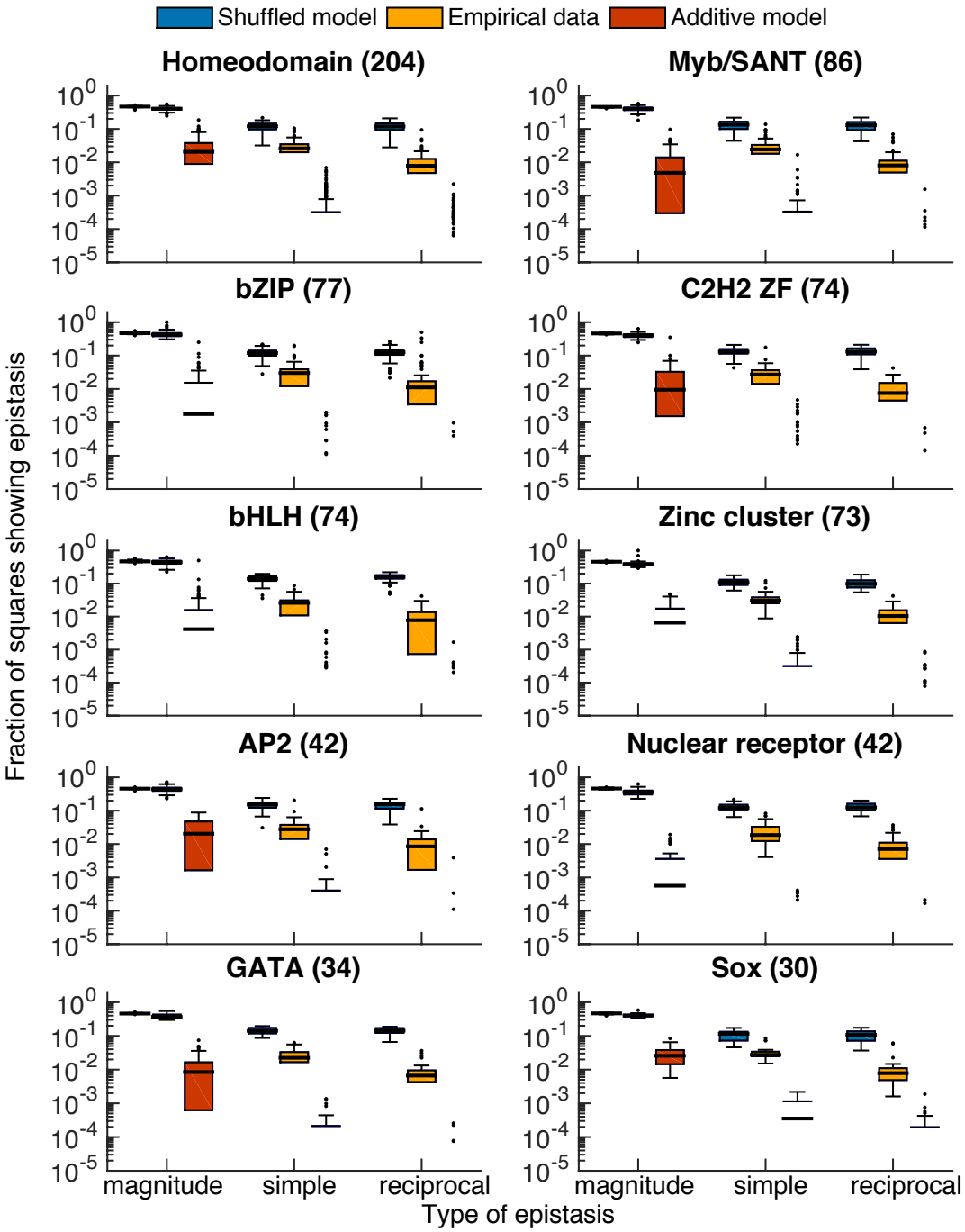
Supplementary Figure 19. Accessibility of the highest-affinity site increases as the binding affinity threshold increases. The data shown represent a sensitivity analysis of the trends presented in Fig. 2C for mutational paths of length 4, which are the most abundant in our dataset. Data points depict the mean peak accessibility in the empirical data and the two null models (see legend), as a function of the binding affinity threshold τ. Error bars depict one standard deviation.



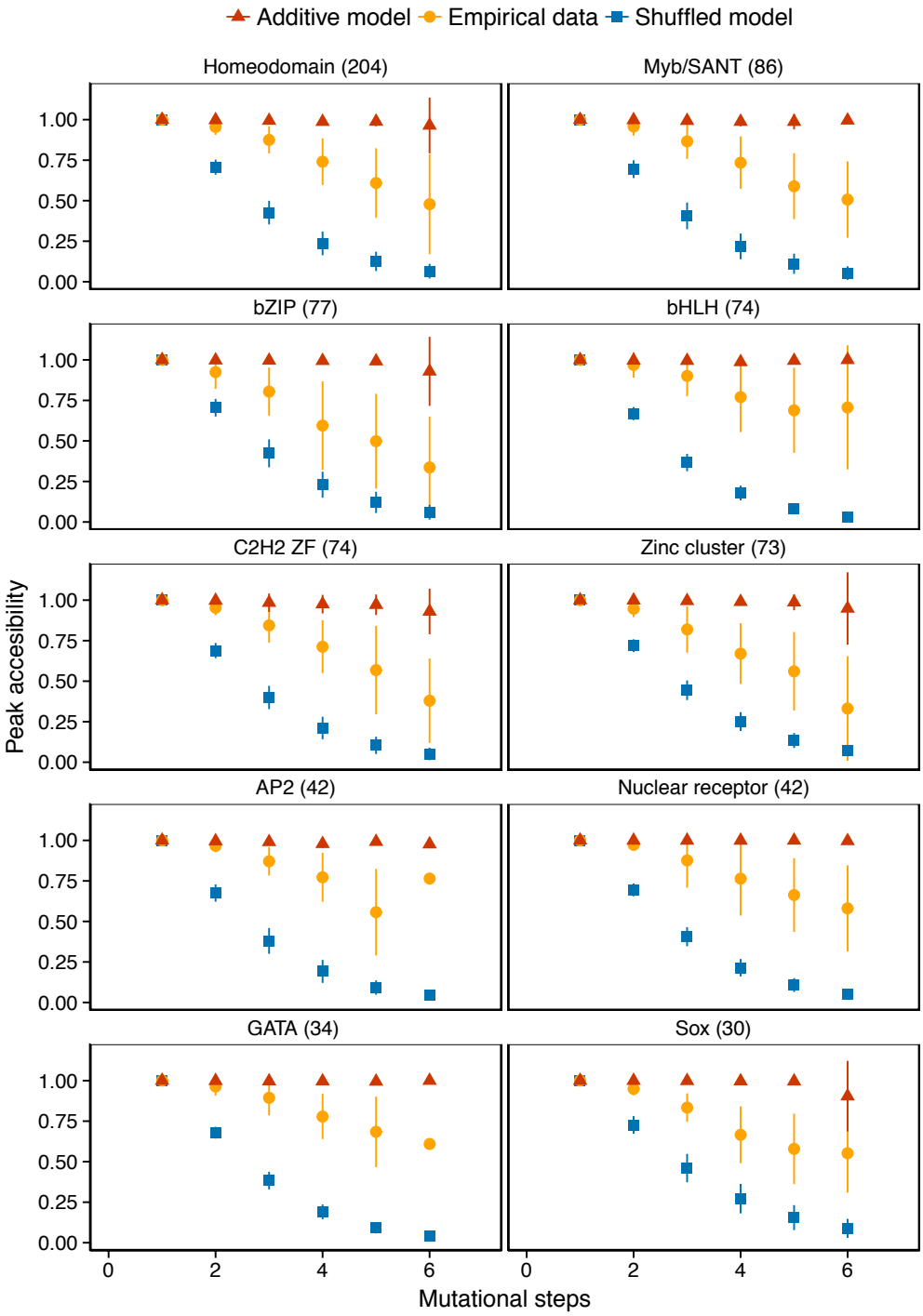
Supplementary Figure 20. The number of peaks is intermediate to that of the additive and shuffled models across DNA binding domains. These data represent a sensitivity analysis of the trends presented in Fig. 2A. Each panel shows the distribution of the number of peaks per landscape for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to a different DNA binding domain family, chosen because they are the five most prominent in our dataset.



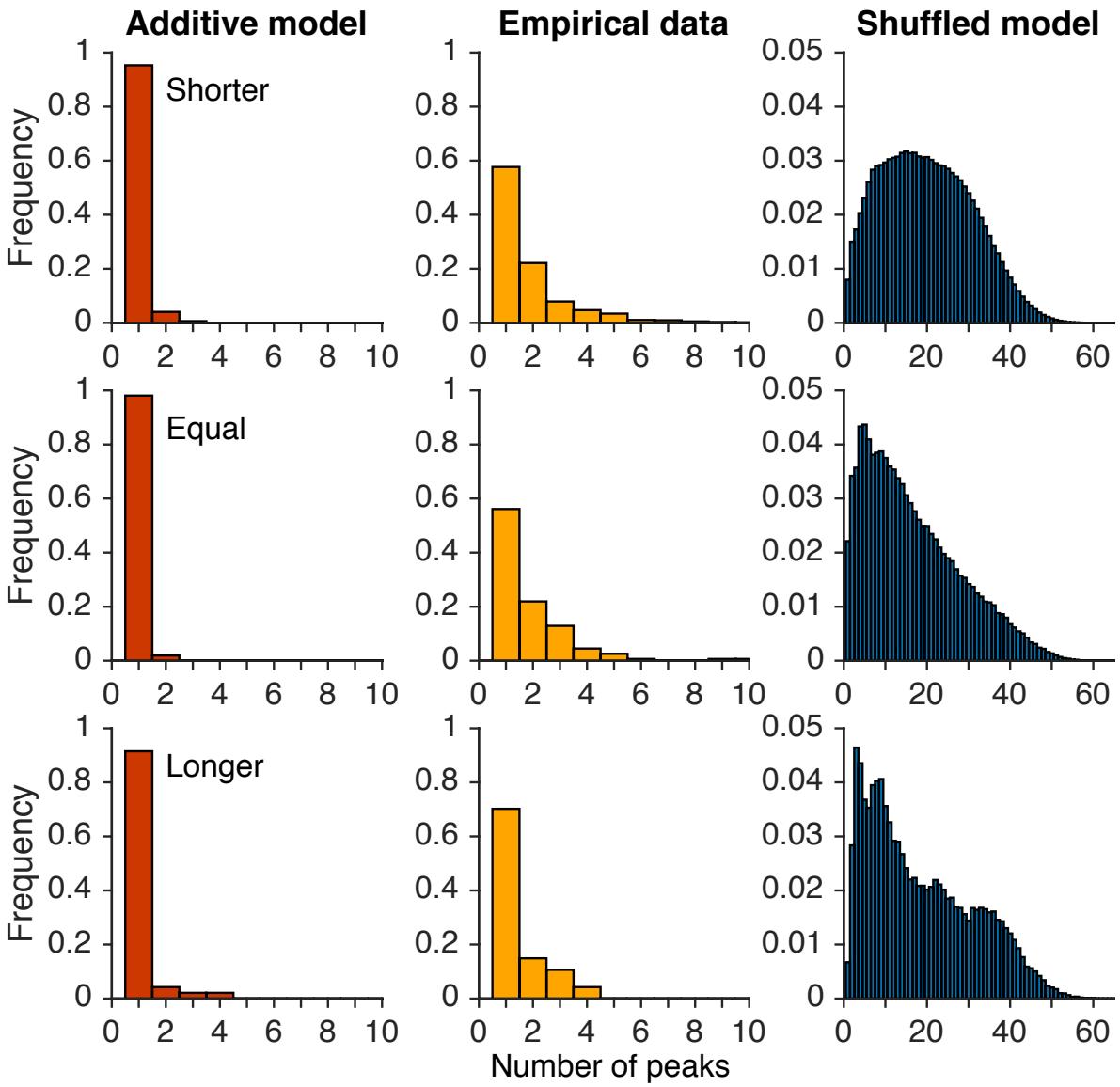
Supplementary Figure 21. The size of the global peak is intermediate to that of the additive and shuffled models across DNA binding domains. These data represent a sensitivity analysis of the trends presented in Supplementary Fig. 7. Each panel shows the distribution of the number of binding sites in the highest peak (i.e. global peak breadth) for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to a different DNA binding domain family.



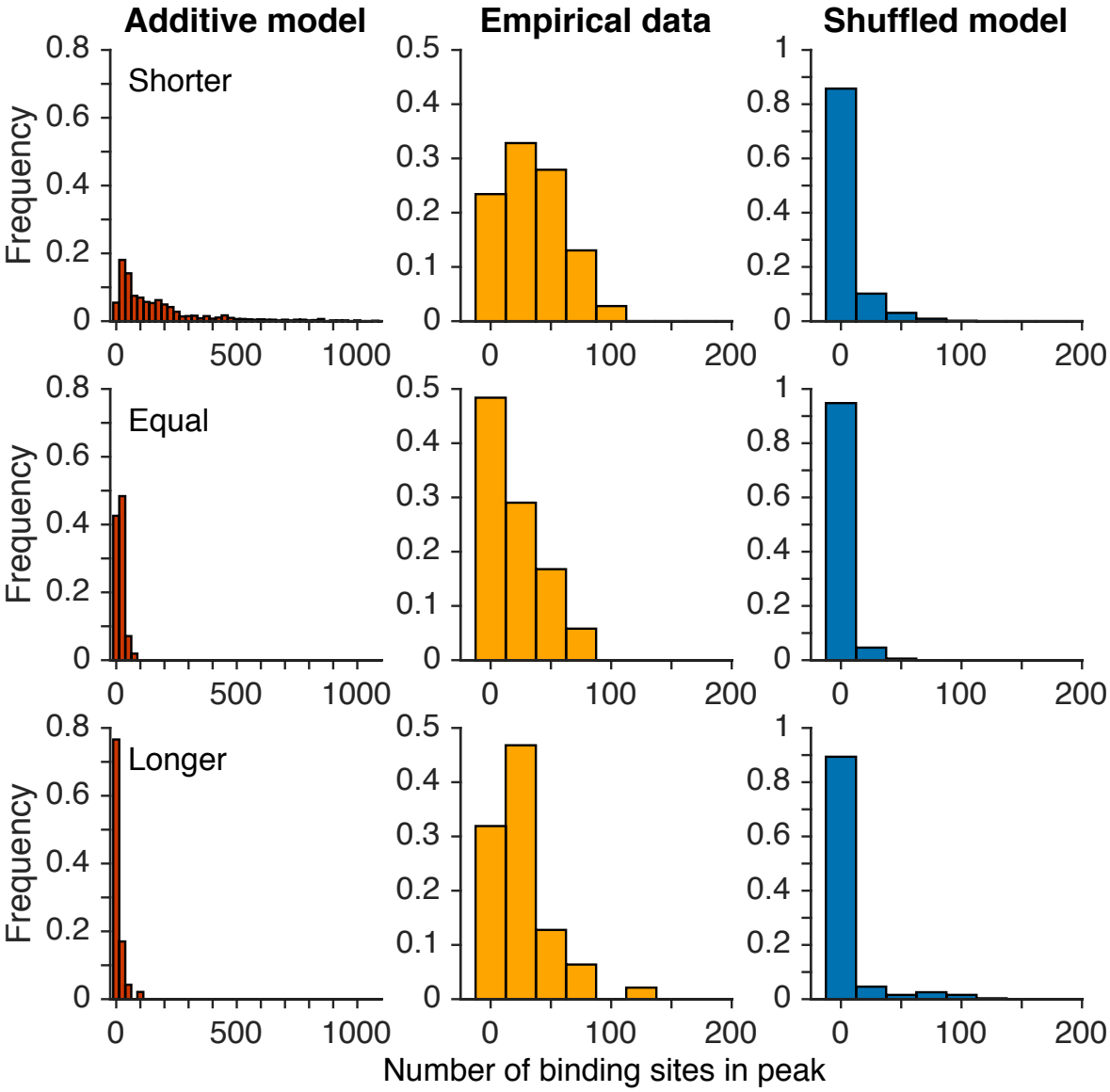
Supplementary Figure 22. Epistasis in the empirical data is intermediate to that of the additive and shuffled models, irrespective of DNA binding domain. These data represent a sensitivity analysis of the trends presented in Fig. 2B. Boxplots of magnitude epistasis, simple sign epistasis, and reciprocal sign epistasis for TFs from the 10 most prominent DNA binding domains in our dataset. The number of TFs in our dataset per DNA binding domain is shown in parentheses. Note the logarithmic scale of the y-axes. Note the logarithmic scale of the y-axes, which obscures data points with zero epistasis. The absence of the thick horizontal line indicates that the median of the distribution is below the lowest value of the y-axis.



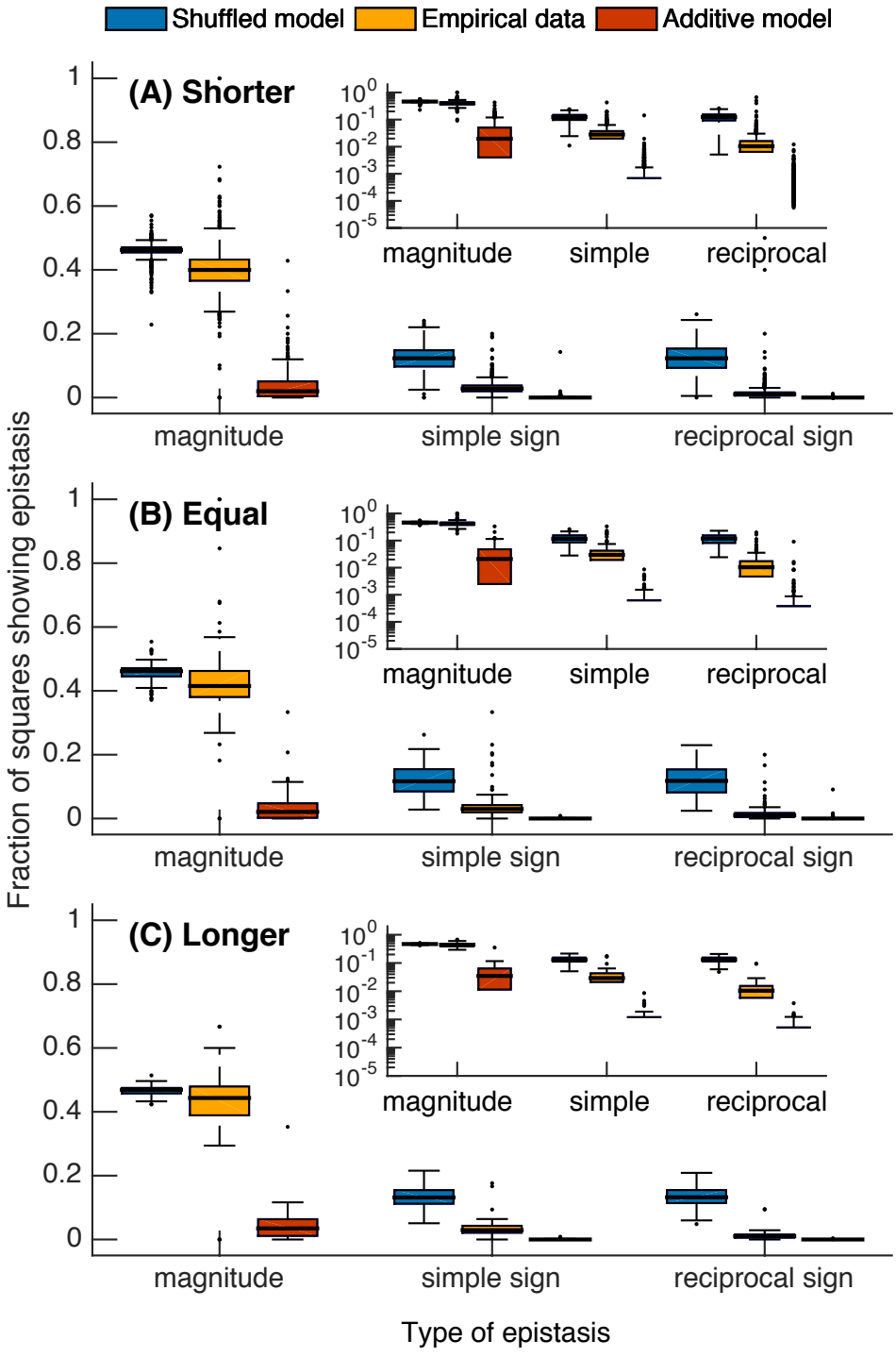
Supplementary Figure 23. Peak accessibility in the empirical data is intermediate to that of the additive and shuffled models, irrespective of DNA binding domain. These data represent a sensitivity analysis of the trends presented in Fig. 2C. Data points depict the mean peak accessibility in the empirical data and the two null models (see legend), as a function of the mutational distance, for the 10 most prominent DNA binding domains in our dataset. The number of TFs in our dataset per DNA binding domain is shown in parentheses. Error bars depict one standard deviation.



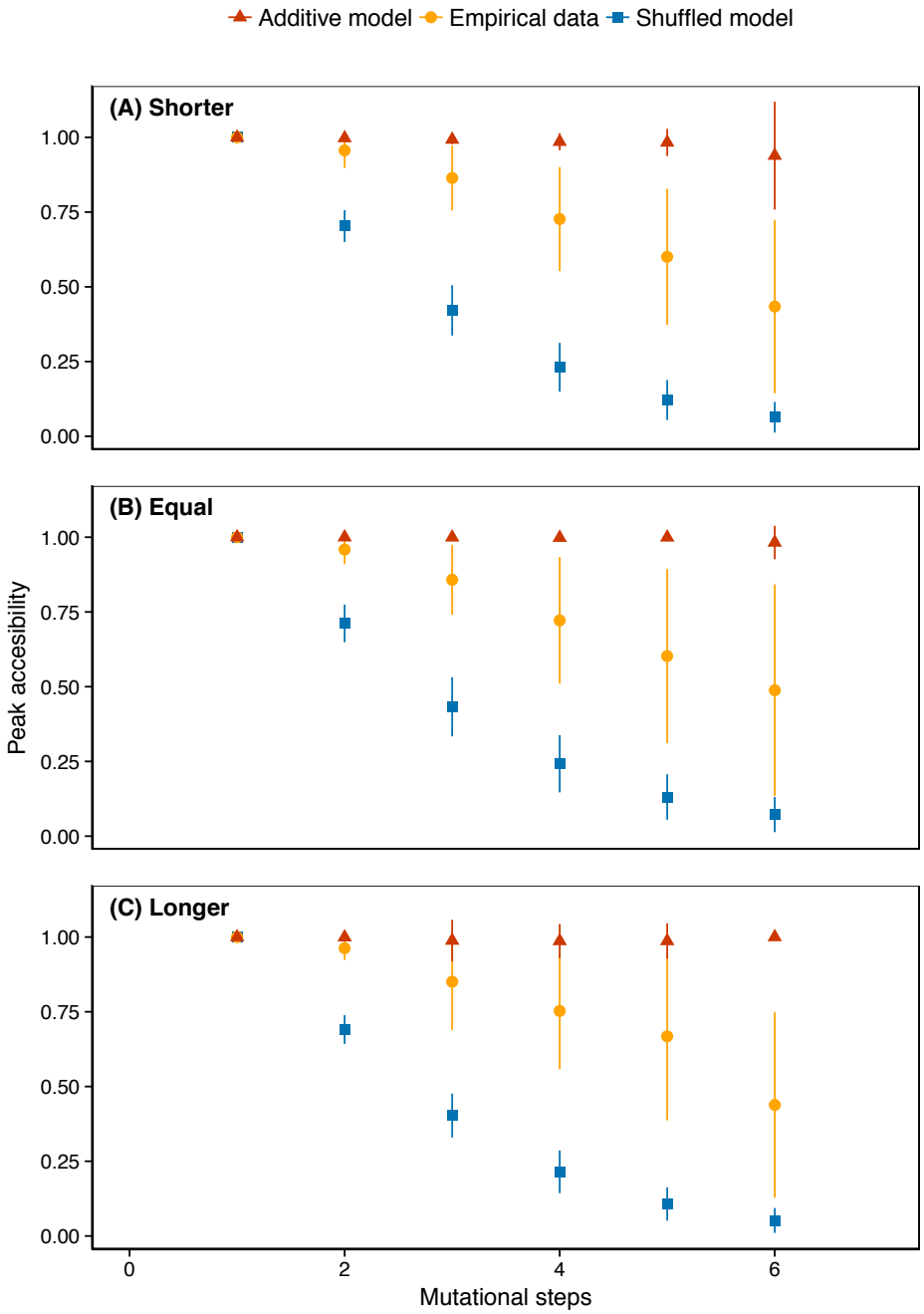
Supplementary Figure 24. The number of peaks in the empirical data is intermediate to that of the additive and shuffled models for TFs that bind sequences that are shorter or longer than eight nucleotides. These data represent a sensitivity analysis of the trends presented in Fig. 2A. Each panel shows the distribution of the number of peaks per landscape for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to TFs that bind sequences that are shorter, equal to, or longer than 8 nucleotides.



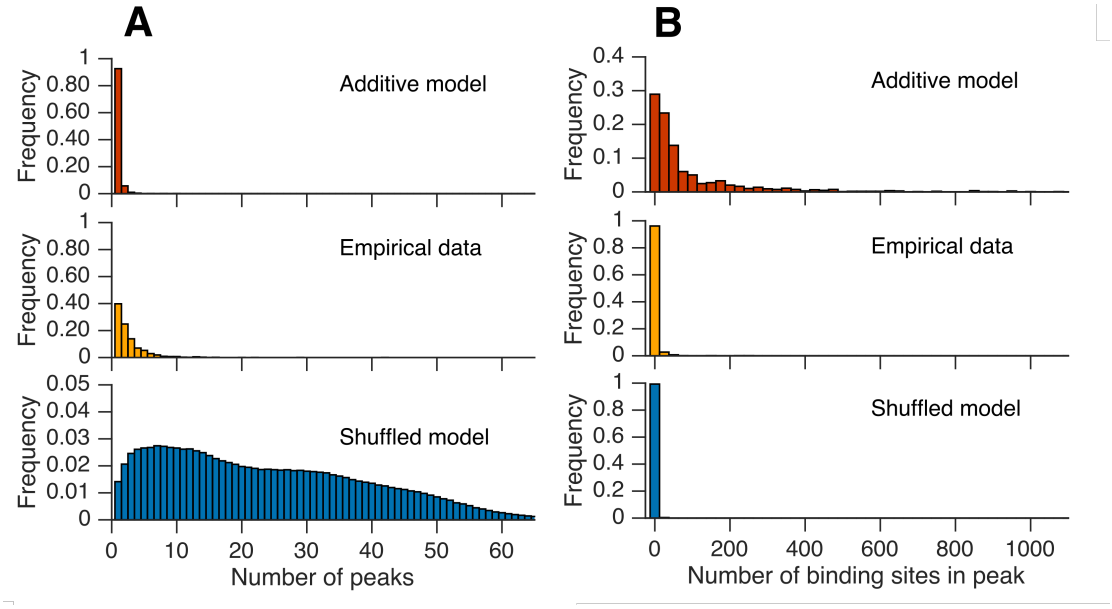
Supplementary Figure 25. The global peak breadth in the empirical data is intermediate to that of the additive and shuffled models for TFs that bind sequences that are shorter or longer than eight nucleotides. These data represent a sensitivity analysis of the trends presented in Supplementary Fig. 7. Each panel shows the distribution of the number of binding sites in the highest peak (i.e. global peak breadth) for the additive model (left column), empirical data (middle column), and shuffled model (right column). Each row of panels corresponds to TFs that bind sequences that are shorter, equal to, or longer than 8 nucleotides.



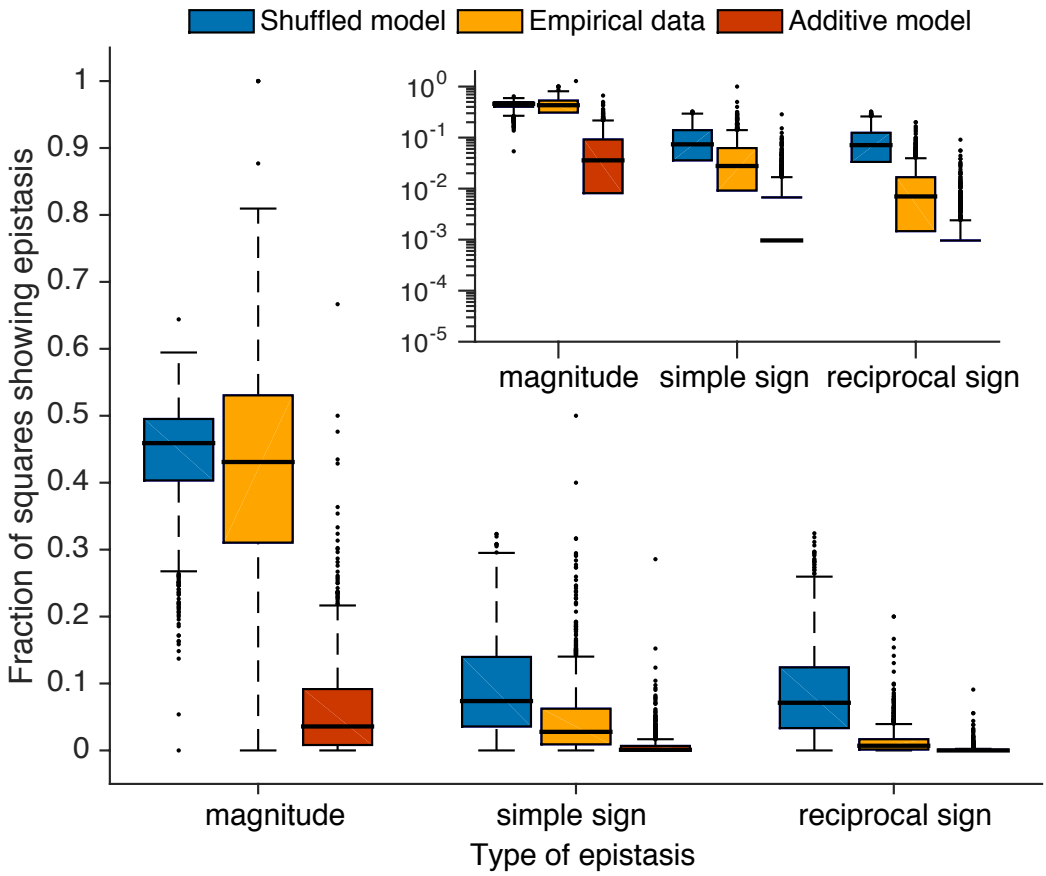
Supplementary Figure 26. Epistasis in the empirical data is intermediate to that of the additive and shuffled models for TFs that bind sequences that are shorter or longer than eight nucleotides. These data represent a sensitivity analysis of the trends presented in Fig. 2B. Boxplots of magnitude epistasis, simple sign epistasis, and reciprocal sign epistasis for TFs that bind sequences that are (**A**) shorter, (**B**) equal to, or (**C**) longer than 8 nucleotides. Note the logarithmic scale of the y-axes, which obscures data points with zero epistasis. The absence of the thick horizontal line indicates that the median of the distribution is below the lowest value of the y-axis.



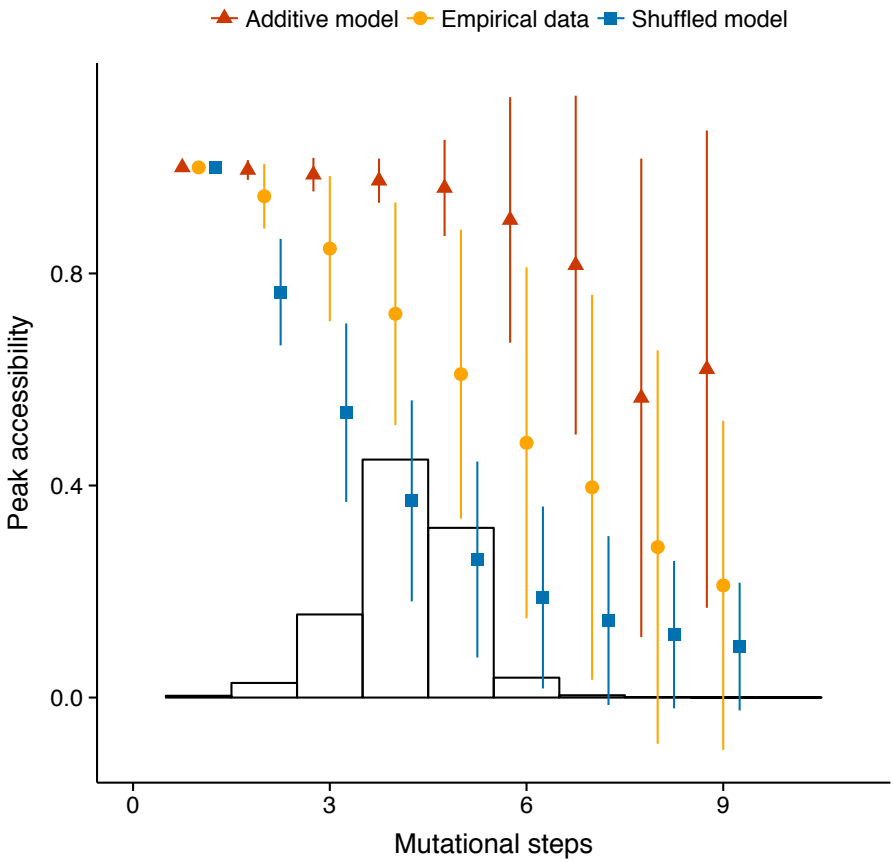
Supplementary Figure 27. Peak accessibility in the empirical data is intermediate to that of the additive and shuffled models for TFs that bind sequences that are shorter or longer than eight nucleotides. These data represent a sensitivity analysis of the trends presented in Fig. 2C. Data points depict the mean peak accessibility in the empirical data and the two null models (see legend), as a function of the mutational distance to the peak sequence, for TFs that bind sequences that are (**A**) shorter, (**B**) equal to, or (**C**) longer than 8 nucleotides. Error bars depict one standard deviation.



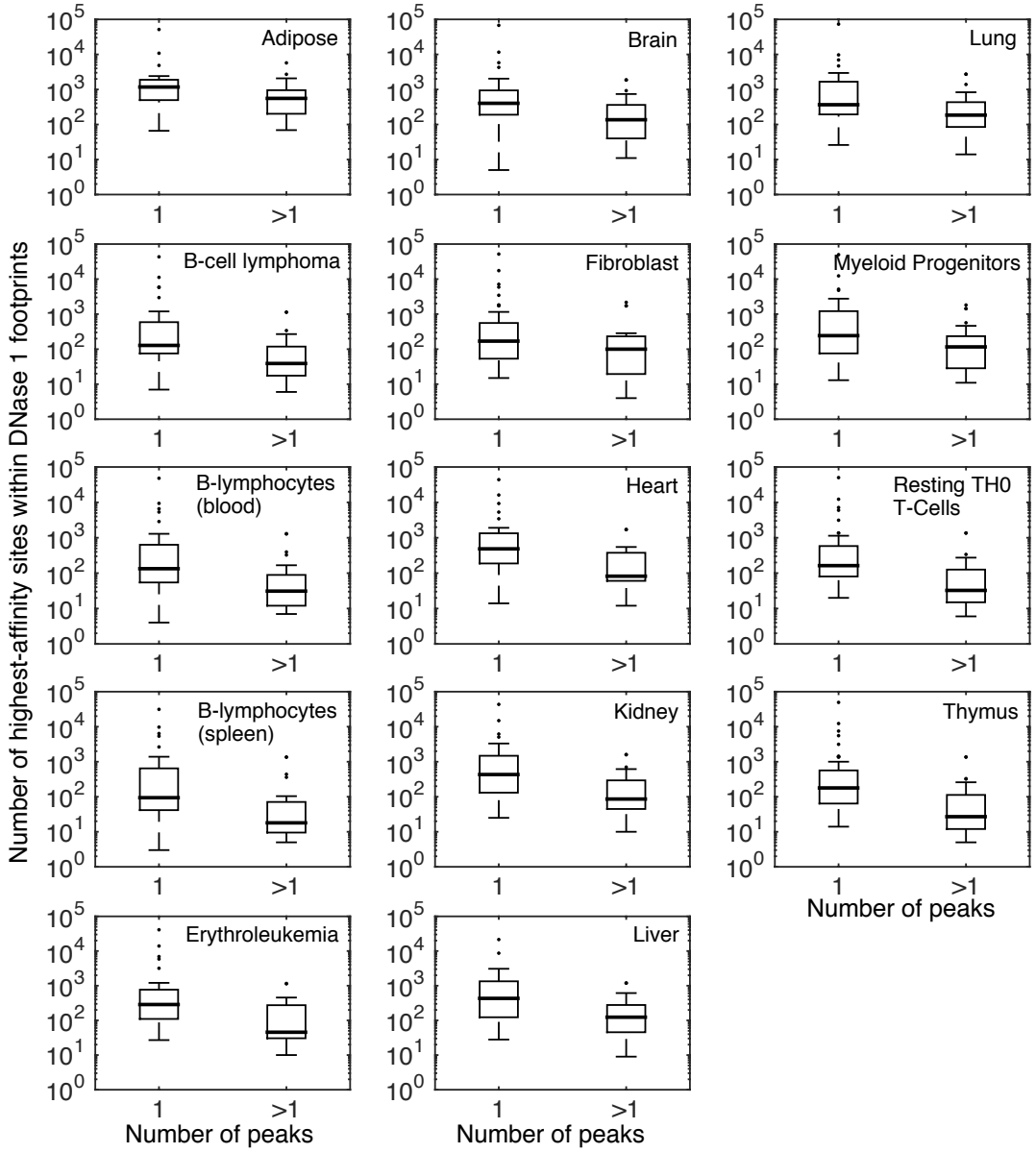
Supplementary Figure 28. The number of peaks in the empirical data is intermediate to that of the additive and shuffled models when using *Z*-scores as a proxy for binding affinity. The distribution of (**A**) the number of peaks, and (**B**) the number of binding sites per global peak for 1,095 empirical adaptive landscapes (yellow) and the additive (red) and rugged (blue) null models. These data represent a sensitivity analysis of the trends presented in Fig. 2C and Supplementary Fig. 7, respectively.



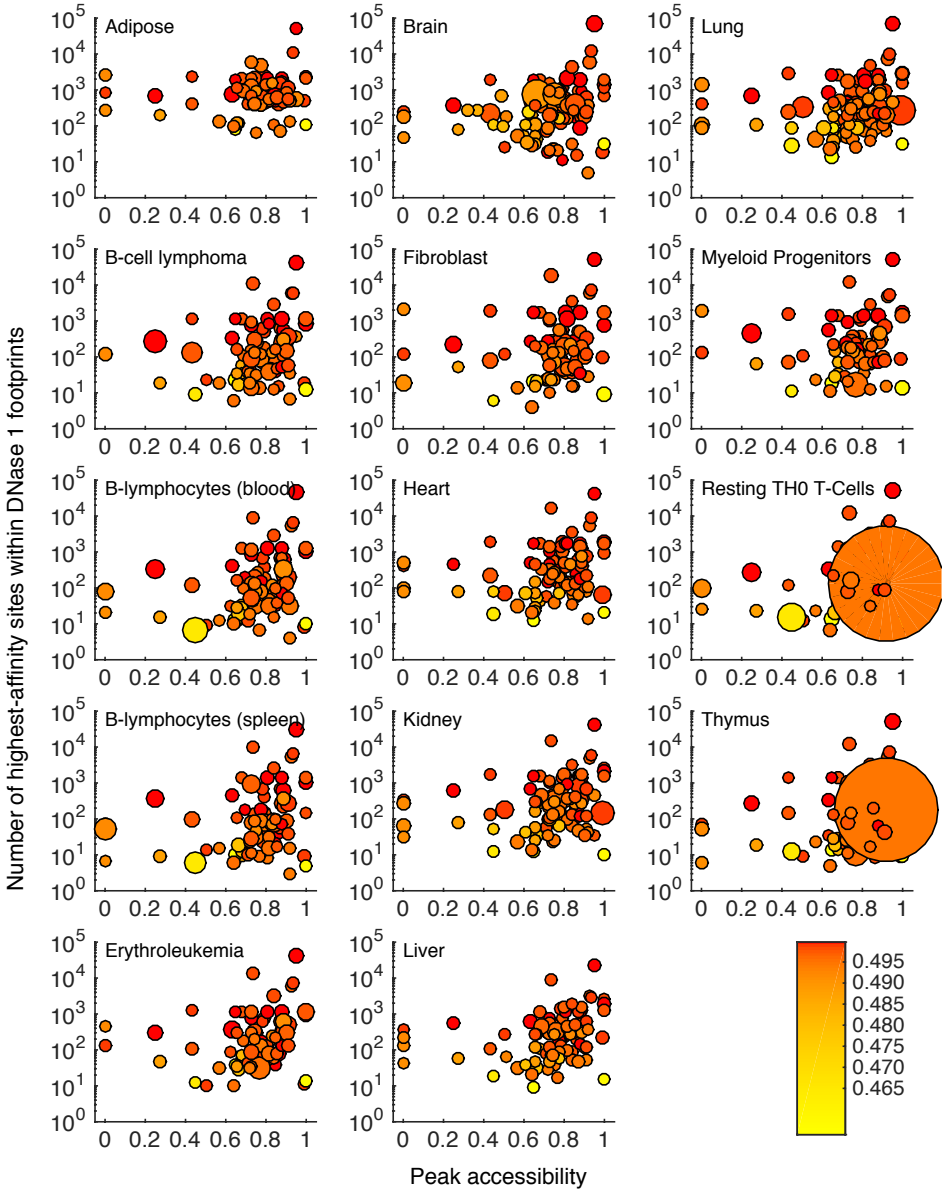
Supplementary Figure 29. Epistasis in adaptive landscapes of transcription factor binding affinity when using *Z*-scores as a proxy for binding affinity. These data represent a sensitivity analysis of the trends presented in Fig. 2B. Boxplots of the fraction of squares showing magnitude, simple sign, and reciprocal sign epistasis for 1,095 adaptive landscapes. The thick horizontal line in the middle of each box represents the median of the data, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. For the shuffled model, the boxplot summarizes 1,095 data points, each of which is an average over 1,000 shuffled landscapes. The inset shows the same data, but with a logarithmically scaled y-axis. The absence of the thick line indicates that the median of the distribution is below the lowest value of the logarithmically scaled y-axis.



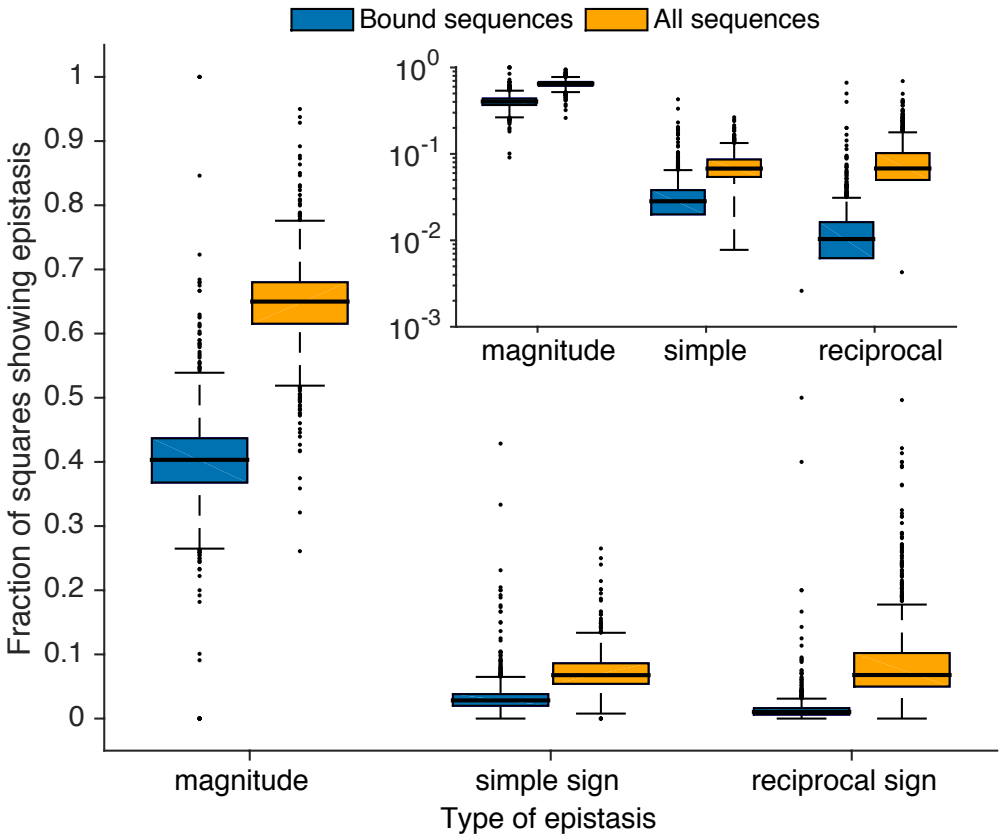
Supplementary Figure 30. Peak accessibility when using *Z*-scores as proxy for binding affinity. These data represent a sensitivity analysis of the trends presented in Fig. 2C. For each of the 1,095 adaptive landscapes we show the mean (symbols) and standard deviation (error bars) of the fraction of accessible paths to the highest-affinity site in the landscape (i.e., peak accessibility). The histogram shows the distribution of mutational steps from the highest-affinity site for all sequences in all landscapes. For the shuffled model, each symbol represents the mean of 1,095 data points, each of which is an average over 1,000 shuffled landscapes.



Supplementary Figure 31. *In vivo* binding site abundance is higher for TFs with single-peaked landscapes than for TFs with multi-peaked landscapes. The vertical axis of each panel indicates the abundance of a TF's highest-affinity site in protein-bound regions of the *M. musculus* genome in 14 cell and tissue types (according to DNase I footprint data; Materials and Methods). The horizontal axis indicates the number of peaks. Landscapes are classified into two categories: single-peaked and multi-peaked. The thick horizontal line in the middle of each box represents the median of the data, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. Each panel corresponds to a different cell or tissue type. Note the logarithmic scale of the y-axis for all the panels. Results of Wilcoxon rank-sum tests can be found in Supplementary Table 3.



Supplementary Figure 32. *In vivo* binding site abundance correlates with peak accessibility. The vertical axis of each panel indicates the abundance of a TF's highest-affinity site in protein-bound regions of the *M. musculus* genome in 14 cell and tissue types (according to DNase I footprint data; Materials and Methods). The horizontal axes show peak accessibility through mutational paths of length 4, which are the most abundant paths in our dataset (Fig. 2C). Each panel corresponds to a different cell and tissue type, and each circle corresponds to a single TF that is expressed in that cell or tissue type. The colour of each circle indicates the binding affinity of the TF's highest-affinity site (darker = higher; colour bar). The size of a circle corresponds to the TF's expression level (larger = higher, Materials and Methods). Note the logarithmic scale of the y-axis for all the panels. The Spearman's rank correlation coefficients and *P*-values can be found in Supplementary Table 3.



Supplementary Figure 33. Our measures of epistasis based only on bound sequences are conservative. Box plots of the fraction of squares showing magnitude, simple sign, and reciprocal sign epistasis for 1,137 adaptive landscapes. The thick horizontal line in the middle of each box represents the median of the data, while the bottom and top of each box represent the 25th and 75th percentiles, respectively. The inset shows the same data, but with a logarithmically scaled y-axis. Bound sequences are binding sites whose binding affinity is above an affinity threshold τ of 0.35 (Materials and Methods).

2. Supplementary tables

Supplementary Table 1. Information about the 1,137 transcription factors and their landscapes.

Supplementary Table 2. Sources for the DNase I hypersensitive regions, footprints, and RNA-seq data for the 14 cell and tissue types analysed in this study.

Supplementary Table 3. Spearman's rank correlations between peak accessibility and the abundance of the highest-affinity site in DNase I footprints for the 14 cell and tissue types, in addition to Wilcoxon rank-sum tests comparing binding site abundance amongst single-peaked and multi-peaked landscapes.

Supplementary Table 4. Spearman's rank correlations between peak accessibility and the abundance of the highest-affinity site in DNase hypersensitive regions for 14 murine cell and tissue types, after having removed the DNase I footprints, in addition to Wilcoxon rank-sum tests comparing binding site abundance in the same hypersensitive regions amongst single-peaked and multi-peaked landscapes, This is a negative control.

Supplementary Table 5. Analysis of covariance (ANCOVA) and its underlying assumptions. The best linear models (those with a higher coefficient of determination, R^2) were obtained when the site abundance and information content were transformed logarithmically.

Supplementary Table 6. Partial Spearman's rank correlations with binding affinity between peak accessibility and the abundance of the highest-affinity site in DNase I footprints for the 14 cell and tissue types.

Supplementary Table 7. Partial Spearman's rank correlations with PWM information content between peak accessibility and the abundance of the highest-affinity site in DNase I footprints for 14 murine cell and tissue types.

3. Supplementary results

3.1. Summary statistics of genotype networks

Here, we summarize some of the structural properties of the genotype networks that serve as the substrate of the adaptive landscapes we study. We observe that depending on the TF, between 9 and 1,186 sites are bound (Supplementary Table 1). For 99.6% of the TFs (1,132 of 1,137), the majority of bound sites form a single connected component in this network, which we refer to as the *dominant* network, and for 53% of the TFs (600 of 1,137, Supplementary Fig. 1A), all of the bound sites belong to this dominant network (corroborating our previous findings based on fewer TFs[24]). For the vast majority of TFs, the vast majority of bound sequences are in the dominant genotype network (Supplementary Fig. 2), while many of the non-dominant networks comprise very few sequences, and predominantly only a single sequence (Supplementary Fig. 3). Therefore, we carry out all of our analyses on the dominant genotype networks, which contain on average 392 sites but vary broadly in size among the TFs (Supplementary Fig. 1B).

3.2. Global peaks are usually organized into broad plateaus

We find that global peaks in binding affinity landscapes rarely comprise only a single sequence; rather they are plateaus typically made up of dozens to hundreds of sequences. The average plateau is large and comprises 10.4% of all binding sites in the landscape (Supplementary Fig. 7), many more than expected in the shuffled model, but far fewer than expected in the additive model. While larger landscapes also tend to have larger plateaus (Spearman's rank correlation coefficient = 0.7, P-value < 2.2×10^{-16}), the fraction of a landscape's binding sites falling on a plateau decreases with landscape size (Supplementary Fig. 6B; Spearman's rank correlation coefficient = -0.35, P-value < 2.2×10^{-16}).

This observation of global peaks being organized into broad plateaus is in line with genome-wide assays of TF binding, which often find non-consensus sequences in bound regions of the genome[43,65]. We stress, however, that our measure of peak breadth is sensitive to the use of E-scores (rather than Z-scores) as the quantitative phenotype in our landscapes (Supplementary section 3.6.5), and this sensitivity should be taken into account when interpreting our findings.

3.3. Why epistasis occasionally appears in the additive null model

Epistasis appears at very low frequencies in the additive model. For example, an average of 0.036% of squares per genotype network exhibit simple sign epistasis, and 0.007% exhibit reciprocal sign epistasis. On the rare occasion where epistasis appears in the additive model, it stems directly from our sliding-window approach for scoring sequences (Materials and Methods). Supplementary Fig. 8 shows an example. The mouse TF Arid5a has a PWM that is 14 nucleotides wide (Supplementary Fig. 8A,B). The mutational pair `AATTTTAA` and `AAATATAA` exhibits reciprocal sign epistasis (Supplementary Fig. 8C), since both mutational intermediates have lower binding affinity. This results from the fact that the highest score for the sequence `AATTTTAA` occurs when it is aligned to positions 3 through 10 in the PWM, whereas the highest score for the other three sequences in the square occur when they are aligned to positions 2 through 9. If `AATTTTAA` were also aligned to positions 2 through 9, then this square would not exhibit epistasis.

While our sliding-window approach is therefore responsible for a very low incidence of epistasis in the additive null model, we believe that it is the most sensible approach for scoring sequences, for at least two reasons. First, most of the PWMs in our dataset are not eight nucleotides wide, and in these cases it is simply necessary to use a sliding window. Second, even for the PWMs that are eight nucleotides wide, the sliding window is superior to a fixed window. To understand why, consider a hypothetical example where the information content in each position of the PWM is maximal, i.e., there is no variation in the nucleotide frequencies per position. Such a PWM could be represented by a single consensus sequence, say `TATATATA`. Using a fixed window, the sequence `ATATATAT` would not match the PWM in a single position and the sequence would receive a score of zero. However, by sliding the sequence by only one position, it will receive a high score, as it should, since seven of its eight nucleotides exactly match the consensus.

3.4. Sign epistasis preferentially occurs among nucleotides that are near one another in the binding site

The prevalence of epistatic interactions in the TF binding sites of 411 mouse and human TFs was recently investigated using data from high-throughput SELEX[30]. For each TF, pairwise epistasis was

estimated as the ratio of the observed frequency of a mutational pair in a binding site to its expected frequency, which was calculated using the TF's PWM. This ratio therefore describes a deviation from additivity. Based on this analysis, the authors concluded that the individual nucleotides of a binding site generally contribute additively to binding affinity, and that when epistatic interactions do occur, they occur preferentially among nucleotides that are near one another in the binding site.

Our analyses generally agree with these observations, but we also extend these earlier findings by partitioning epistasis into three distinct classes that have increasingly detrimental effects on landscape navigability: magnitude epistasis, simple sign epistasis, and reciprocal sign epistasis (Box 1). Specifically, we find that both simple sign epistasis (Supplementary Fig. 9B) and reciprocal sign epistasis (Supplementary Fig. 9C) decrease as the distance between the binding site positions increases, but we observe no such trend for magnitude epistasis (Supplementary Fig. 9A). Thus, the two classes of epistasis that hamper landscape navigability the most – simple and reciprocal sign epistasis – occur preferentially among nearby binding site positions, whereas the class of epistasis that does not affect landscape navigability – magnitude epistasis – has no such preference.

3.5. Peak accessibility decreases when unbound sequences are included

As with our measure of epistasis (Supplementary section 3.8), we restricted our measure of peak accessibility to bound sequences. Including unbound sequences can only decrease the navigability of a landscape, and therefore decrease peak accessibility. To determine the extent of this decrease, we repeated our peak accessibility analyses for mutational paths that include unbound sequences. Supplementary Fig. 11 shows that the measures of peak accessibility reported in the main text are always higher than those that include unbound sequences (especially at longer mutational distances), as expected. The reason is twofold. First, since mutational paths always start at a bound sequence (i.e., $\tau > 0.35$), any path that includes an unbound sequence (i.e., $\tau \leq 0.35$) is by definition inaccessible. Second, the inclusion of unbound sequences creates new shortest paths to the global peak that are inaccessible and that occasionally supplant longer, accessible shortest paths to the global peak that were present in the genotype networks studied in the main text.

3.6. Sensitivity analyses

Our study employs two parameters, one for determining whether two DNA sequences differ in their binding affinity (δ) and another to determine which DNA sequences are bound by a TF (τ). The threshold δ is important because it accounts for the experimental noise in PBM data[52]. Increasing δ has the effect of "smoothing" the landscapes (Fig. **1**B), because larger values of δ make it less likely that two sequences have different binding affinities. We use a unique threshold δ for each TF, which is derived from noise in experimental and replicated PBM measurements for that particular TF (Materials and Methods). Increasing the affinity threshold τ reduces the number of bound sequences, and decreases the size of a landscape (Fig. **1**C). Below, we assess the sensitivity of our main results to broadly varying parameter values for τ and δ. Moreover, we determine whether our main results differ among TFs from distinct DNA binding domain structural classes, as well as among TFs that bind DNA sequences shorter or longer than eight nucleotides. Finally, we show that our measure of global peak breadth is sensitive to the use of *E*-scores (rather than *Z*-scores) as a quantitative phenotype.

Specifically, we test the sensitivity of the following main results: (1) The empirical landscapes have more peaks than expected under the additive model, but far fewer than expected under the shuffled model. (2) The global peaks in the empirical landscapes comprise more sequences than expected under the shuffled model, but fewer than expected under the additive model. (3) The incidence of the three classes of epistasis in the empirical landscapes is higher than expected under the additive model, but lower than expected under the shuffled model. (4) The highest-affinity sites are more accessible in the empirical data than expected under the shuffled model, but less accessible than those expected under the additive model.

3.6.1. Our observations are insensitive to broadly varying thresholds for noise filtering

In the main text, we used a threshold δ to determine whether two sequences differed in their binding affinity. This threshold is important because it accounts for the inherent noise of protein-binding microarray data[52]. Here, we explore the sensitivity of our results to this parameter.

As the noise threshold δ increases, the number of peaks decreases, for the empirical data and both null models (Supplementary Fig. 12). This is because increasing the noise threshold makes it less

likely for two sequences to have different binding affinities, which has the effect of merging peaks. Nonetheless, even for the most permissive noise threshold of $\delta = 0.05$ (which corresponds to one-third of the range in affinity values when $\tau = 0.35$), 32% of the empirical landscapes comprise multiple peaks, as opposed to 3% of the additive landscapes and 98% of the shuffled landscapes.

Supplementary Figure 13 shows that the number of binding sites per global peak (i.e., global peak breadth) increases as the noise threshold δ increases, for the empirical data and both null models. This occurs because increasing δ makes the global peak more "inclusive," such that the binding affinity of a sequence can differ from that of the highest-affinity sequence by a greater amount and the sequence will still be considered part of the global peak.

Supplementary Figure 14 shows that epistasis decreases as the noise threshold δ increases, for all three classes of epistasis, in the empirical data and both null models. This occurs because increasing δ decreases the likelihood that Eqs. (3) or (4) (Materials and Methods) are satisfied, which decreases the proportion of mutational pairs that are classified as epistatic. These sensitivity analyses confirm intuition: increasing the noise threshold δ increases the navigability of an adaptive landscape. This is further evidenced by Supplementary Fig. 15, which shows that peak accessibility increases with δ in both the empirical data and the two null models. With the exception of magnitude epistasis – which does not affect landscape navigability – at low δ, all measurements of epistasis and peak accessibility in the empirical data are intermediate to those of the additive and shuffled null models. We therefore conclude that the adaptive landscapes of TF binding preferences are less navigable than those from the additive model, but far more navigable than landscapes from the shuffled model, and that this observation is insensitive to changes in δ.

3.6.2. Our observations are insensitive to broadly varying affinity thresholds for delineating bound from unbound sequences

In the main text, we considered a sequence "bound" if its *E*-score exceeded 0.35. Here, we assess the sensitivity of each of our measures of landscape navigability to the choice of binding affinity threshold τ.

As the affinity threshold τ increases, the distribution of the number of peaks shifts toward lower values, such that an increasing number of landscapes become single-peaked (Supplementary Fig. 16). This is true for the empirical data and for both null models, and results from the fact that increasing the binding affinity threshold decreases the number of bound sequences[24], thus pruning non-global peaks from the landscape. However, even for the most stringent threshold of $\tau = 0.45$, under which only very few sequences are considered bound[24], 14% of the empirical landscapes comprise multiple peaks, as compared to 2% of the additive landscapes and 58% of the shuffled landscapes.

Supplementary Figure 17 shows that the empirical distribution of the number of binding sites in the global peak (i.e., global peak breadth) is relatively unaffected by changes in the binding affinity threshold τ. This is because the sequences in these peaks typically have *E*-scores that exceed even our most stringent threshold of $\tau = 0.45$. In contrast, the distributions of global peak breadth under the two null models are sensitive to the binding affinity threshold, but in different ways. In the additive model, the global peak breadth decreases as τ increases. This is because the sequences at the base of each peak are gradually removed from the landscape as τ increases, analogous to the loss of exposed landmass as water levels rise. In the shuffled model, in contrast, the global peak breadth increases as τ increases, a counterintuitive observation that stems from the way these landscapes are constructed. Specifically, increasing τ decreases the number of bound sequences[24], and in a landscape with few sequences, a random permutation of binding affinities is likely to clump many of the high-affinity sequences near one another in the landscape, such that they are part of the same peak. In contrast, in a landscape with many sequences (low τ), these same high-affinity sequences are likely to be spread throughout the landscape, and thus to be part of different peaks. Despite these changes in the landscapes generated by the two null models, the global peaks in the empirical landscapes persist in comprising more sequences than expected under the shuffled model, but fewer than expected under the additive model. We therefore conclude that these observations are insensitive to changes in τ.

Supplementary Figure 18 shows boxplots of the three classes of epistasis as a function of the binding affinity threshold τ, for the empirical data and the two null models. In the additive model, there is very little epistasis and it is generally insensitive to changes in τ. In the empirical data and in the shuffled model, the incidence of epistasis tends to decrease as τ increases, indicating an increase in

landscape navigability. These observations are reflected in Supplementary Fig. 19: peak accessibility increases as the binding affinity threshold τ increases for both the empirical data and the shuffled model, but for the additive model, peak accessibility does not change with τ. Importantly, all measurements of epistasis and peak accessibility in the empirical data are intermediate to those of the additive and shuffled null models. We therefore conclude that the adaptive landscapes of TF binding affinity are less navigable than those from the additive model, but far more navigable than landscapes from a shuffled model, and that this observation is insensitive to changes in τ.

3.6.3. Our observations are consistent across DNA binding domains

The 1,137 TFs in our dataset represent 62 different DNA binding domain structural classes (Supplementary Table 1). Here, we show that our results do not vary among these classes. While we illustrate this point using data from the most prominent DNA binding domains in our dataset, the insensitivity of our results applies to the other structural classes as well (data not shown).

Supplementary Figure 20 shows the distribution of the number of peaks in the empirical data and the two null models for the five most prominent DNA binding domain classes. While the distributions change quantitatively, especially under the shuffled model, they do not change qualitatively. Specifically, the empirical landscapes always have more peaks than expected under the additive model, but fewer than expected under the shuffled model.

Supplementary Figure 21 shows the number of binding sites in the global peak (i.e., global peak breadth) in the empirical data and the two null models, for the same five DNA binding domain structural classes. Again, the distributions do not change qualitatively: Global peak breadth is larger in the empirical landscapes than expected under the shuffled model, but smaller than expected under the additive model.

Supplementary Figure 22 shows that the incidence of the three classes of epistasis among the empirical data and the two null models does not vary among DNA binding domain structural classes. Epistasis in the empirical data is always intermediate to that of the additive and shuffled null models. As a consequence, peak accessibility in the empirical data is intermediate to that of the additive and shuffled null models (Supplementary Fig. 23). We therefore conclude that the adaptive landscapes of

TF binding preferences are less navigable than those from the additive model, but far more navigable than landscapes from the shuffled model, and that this observation does not vary among the DNA binding domain structural classes considered here.

3.6.4. Our observations are consistent across TFs that bind shorter or longer sequences than eight nucleotides

Protein-binding microarray data characterize the binding affinity of a TF to all possible binding sites of length 8. However, many of the TFs in our dataset bind sequences that are shorter or longer than 8 nucleotides. It is therefore important to ensure that our results are insensitive to binding site length. To determine the length of a TF's binding sites, we use the TF's position weight matrix to calculate the maximum number of contiguous nucleotides with information content (Eq. 8, Materials and Methods) exceeding 0.5 bits (bold font in Supplementary Fig. 8B).

We find that among the 1,137 TFs studied here, 935 bind sequences that are shorter than 8 nucleotides and 47 bind sequences that are longer than 8 nucleotides; the remaining 155 TFs bind sequences that are exactly 8 nucleotides long. Supplementary Figure 24 shows that the empirical landscapes comprise more peaks than expected under the additive model, but far fewer than expected under the shuffled model, regardless of binding site length. Supplementary Figure 25 shows that global peaks in the empirical landscapes typically comprise multiple sequences. Notably, for TFs that bind sequences shorter than 8 nucleotides, the distribution of global peak breadth shifts toward higher values under the additive model. This results from our scoring strategy, which assigns a score to each sliding subsequence, and then assigns the maximum of these scores to the sequence (Materials and Methods). Since many sequences will contain the same high-scoring subsequence, global peak breadth increases for TFs that bind sequences shorter than 8 nucleotides under the additive model.

Supplementary Figure 26 shows that the incidence of the three classes of epistasis among the empirical data and the two null models does not vary among TFs that bind sequences that are shorter or longer than 8 nucleotides. Epistasis in the empirical data is greater than expected under the additive model, but less than expected under the shuffled model. These trends are reflected in Supplementary Fig. 27, which depicts peak accessibility as a function of the mutational distance from the peak

sequence. Peaks in the empirical adaptive landscapes are always less accessible than those from the additive model, but much more accessible than those from the shuffled model. We therefore conclude that the adaptive landscapes of TF binding affinity are less navigable than those from the additive model, but far more navigable than landscapes from the shuffled model, and that this observation is insensitive to binding site length.

3.6.5. Peak breadth is sensitive to the use of E-scores as a quantitative phenotype

Protein-binding microarray data typically include both an *E*-score and a *Z*-score for each DNA sequence. In the main text, we considered the *E*-score as the quantitative phenotype that defines the surface of each adaptive landscape. The *E*-score has the advantage that it is robust to outliers, insensitive to TF concentrations, and less variable across arrays. However, it has the disadvantage of compressing the dynamic range of binding affinities, particularly for *E*-scores near the maximum value of 0.5. Here, we assess the sensitivity of our results to the use of *E*-scores for the 1095 landscape for with we have *Z*-scores (Materials and Methods), revealing that global peak breadth is sensitive to this measure, whereas our measures of landscape navigability are not.

Supplementary Figure 28 shows that the number of peaks in the empirical landscapes remains intermediate to that of the additive and shuffled models, although the number of peaks does increase slightly. However, the number of sequences per peak decreases significantly, relative to the landscapes constructed using *E*-scores. For example, the average number of sequences per peak decreases from 33.2 to 3.6, and the number of single-sequence peaks increases from 36 to 650. Thus, our finding that peaks typically comprise dozens to hundreds of sequences should be considered in the context of this sensitivity. In contrast, the three classes of epistasis that we study are insensitive to the use of *E*-scores (Supplementary Fig. 29), as are our measures of peak accessibility (Supplementary Fig. 30).

3.7. The *in vivo* relationship between landscape navigability and the abundance of binding sites is not driven by binding affinity or by information content

We have previously found that the mouse genome is enriched in high-affinity binding sites both genome-wide and within putative enhancers[24]. Therefore, we checked whether any of the *in vivo*

correlations with the abundance of the highest-affinity site in DNase I footprints could be explained by a confounding relationship with binding affinity. To evaluate that possibility, we tested whether the observed differences in site abundance between single-peak and multi-peaked landscapes are still significant after adjusting for binding affinity. An analysis of covariance (ANCOVA)[66] shows that after controlling for differences in the response variable (site abundance) caused by the covariate (binding affinity), there are still significant differences caused by peak number (Supplementary Table 5). In both brain and MEL cells, the assumption of the homogeneity of slopes among groups – one of the underlying assumptions for the ANCOVA – is violated. However, in brain, the main effect of peak number in a linear model that incorporates the interaction between peak number and binding affinity is significant (t-test, *P*-value = 0.038). A partial correlation analysis shows that the effect of peak accessibility still holds after controlling for binding affinity, except for three tissue types (Supplementary Table 6).

To measure *in vivo* binding site abundance, we scanned DNase I footprints using FIMO[53] and each TF's position weight matrix (Materials and Methods). Since the information content of a TF's position weight matrix may impact the number of "hits" detected in such genomic scans, we performed an additional ANCOVA that controlled for this potential confounding factor. However, this analysis was only necessary for brain and heart tissues since for all other cell or tissue types there was no linear relationship between site abundance and information content (Supplementary Table 5). This analysis shows that the effect of peak number on the abundance of the highest-affinity site in DNase I footprints remains statistically significant after controlling for the information content of each TF's position weight matrix (Supplementary Table 5). A partial correlation analysis shows that the same is true for peak accessibility (Supplementary Table 7).

Thus, we conclude that landscape navigability influences the prevalence of high-affinity TF binding sites in protein-bound regions of the mouse genome, independent of binding affinity or the information content of the TF's position weight matrix.

3.8. Our measures of epistasis for bound sequences are conservative

We have restricted our measures of epistasis to bound sequences (i.e. sequences with a binding

affinity above τ; Materials and Methods) because these form the genotype networks that underlie the adaptive landscapes we study. Our measures of epistasis are therefore conservative, because they do not include unbound sequences, and therefore underestimate the level of ruggedness in a landscape. To determine how conservative our measures of epistasis are, we performed an additional analysis in which we included both bound and unbound sequences. Specifically, in each landscape, we analysed all pairs of bound sequences that differed by two mutations and used them to calculate epistasis for all possible mutational paths between these two sequences (i.e., all squares), regardless of whether the intermediate sequences were part of the genotype network or not. Supplementary Fig. 33 shows that the measures of epistasis reported in the main text are always lower than those that include unbound sequences, as expected.

4. Supplementary discussion

Since Sewall Wright introduced the metaphor of the adaptive landscape in the early 1930s[1,67], it has received considerable attention from theorists interested in understanding how landscape topography affects evolutionary dynamics[68]. In recent years, however, attention has shifted from theoretical landscapes to the analysis of empirical landscapes constructed from experimental data[3]. This shift has been triggered by advances in high-throughput sequencing and chip-based technologies, which have made it possible to assign phenotypes or fitness values to a large number of genotypes, thus bringing the landscape metaphor to life at unprecedented resolution. The study of empirical landscapes is currently a burgeoning area of research[69–73].

However, most of the empirical adaptive landscapes studied to date describe only a minute fraction of the full genotype space, and they are therefore highly incomplete. For example, one of the pioneering studies of empirical adaptive landscapes – and a truly inspiring one – considered antibiotic resistance as a function of all possible combinations of five mutations in the TEM-1 beta-lactamase allele of *Escherichia coli*, and concluded that the landscape is single-peaked[31]. However, a subsequent analysis of a greater number of resistance-conferring mutations in the same allele demonstrated that

this landscape is in fact multi-peaked[74], highlighting that the analysis of incomplete landscapes may produce misleading results[75].

Here, we have studied 1,137 complete adaptive landscapes of transcription factor binding affinity using experimental data from protein binding microarrays. We find that they exhibit little epistasis and contain few peaks, and these peaks are easily accessed via a series of small mutations that only move "uphill." We highlight that these findings go far beyond our previous work on genotype networks of transcription factor binding sites[24], in which we studied a binary phenotype – the molecular capacity of a DNA sequence to bind specific transcription factors – and ignored the quantitative measures of binding affinity that are the focus of this study. Including a quantitative measure of binding affinity transforms "flat" genotype networks into "three-dimensional" landscapes.

We have characterized the incidence of epistasis as a measure of landscape navigability, motivated by a recent debate over the significance of epistatic interactions in TF binding sites[61,62], and more generally over the prevalence of epistasis in molecular evolution[76,77]. While such interactions have been convincingly demonstrated for some TFs[78–80], large-scale analyses of mouse and human TFs suggest that epistasis is the exception, rather than the rule, in TF binding[22,30]. Our results extend and complement these earlier findings, by categorizing epistatic interactions as magnitude, simple sign, or reciprocal sign epistasis. This categorization reveals that both forms of sign epistasis are rare in TF binding sites, but that magnitude epistasis is common. However, there are some notable exceptions. For example, the empirical landscapes of four TFs (Dbp, bzpH, e_gw1.279.28.1, PRKRIR) exhibit the maximum value of magnitude epistasis, something that is rarely seen in the shuffled model. More generally, the empirical landscapes of 171 TFs exhibit more magnitude epistasis than the shuffled landscapes, the empirical landscapes of 25 TFs exhibit more sign epistasis than the shuffled landscapes, and the empirical landscapes of 15 TFs exhibit more reciprocal sign epistasis than the shuffled landscapes. These TFs tend to have small dominant genotype networks (the median number of bound sequences is 103), suggesting that high-specificity TFs tend to bind sequences that exhibit high levels of epistasis. Additionally, while sign epistasis preferentially occurs among bases that are near one another in the binding site – as previously reported for human TFs[30] –

magnitude epistasis shows no such preference: Bases at opposite ends of a binding site are just as likely to exhibit magnitude epistasis as are adjacent bases.

Most of what was previously known about TF binding affinity landscapes came from *in vitro* evolution experiments[81] and biophysical models of TF-DNA interactions[12,13,15,16,35,47,82–86]. Specifically, *in vitro* evolution experiments with the bacterial *lac* repressor have hinted that TF binding affinity landscapes are highly navigable[81], because high affinity binding sites evolve quickly from a small randomized library of oligonucleotides via repeated rounds of mutation and selection. Biophysical models draw similar conclusions, but since they were not based on exhaustive measurements of binding affinity, they necessarily made several assumptions about landscape structure. These include the assumption that single nucleotide changes in a binding site result in small changes in binding affinity[86–88], that a binding site's constituent bases contribute additively to affinity[15,82], and that binding affinity is a linear function of a site's mutational distance from the highest-affinity site[35,86]. We and others have provided broad support for the first assumption, because the binding affinity of a sequence is strongly correlated with that of its mutational neighbors[5,24]. In contrast, the latter assumptions are occasionally violated: Magnitude epistasis is common and landscapes sometimes have multiple peaks. These findings urge caution when incorporating additivity assumptions into models of TF-DNA interactions.

4.1. Caveats

This study has some caveats that are worth highlighting. First, the landscapes we have constructed make several simplifying assumptions about transcriptional regulation. For instance, since they are based on an *in vitro* assay of TF binding, they do not capture many of the complexities of binding *in vivo*, such as epigenetic marks[89], chromatin context[90], local sequence context[91], or interactions with protein partners[42,92], all of which are likely to affect landscape topography. Nevertheless, we observe significant correlations among our measures of landscape navigability and the *in vivo* abundance of high-affinity sites and gene expression levels, suggesting that these *in vitro* affinity landscapes provide meaningful information about TF binding *in vivo*. This is further supported by the observation that the genetic diversity of binding sites across yeast strains reflects the size of their landscape's

global peaks. The second caveat is that we cannot completely rule out the possibility that some of the correlations we observe are spurious. While their consistency across different species, experiments, as well as diverse cell and tissue types is reassuring, only direct experimentation could conclusively determine whether the navigability of a binding affinity landscape affects the evolution of a TF's binding sites *in vivo*. High-throughput laboratory evolution experiments with selection for transcription factor binding could shed light on which landscape parameterizations (τ and δ) best reflect binding *in vivo*, and could provide answers to long-standing questions about the influence of landscape navigability on adaptation, mutational robustness, and the predictability of evolution. Our current research is directed along these lines. Third, the metaphor of the adaptive landscape commonly evokes the fitness of an entire organism, and it is not immediately clear how TF binding affinity relates to organismic fitness. Even though examples – such as the evolution of an activator's binding site in the promoter of an antibiotic resistance gene[10] – have linked binding affinity to the fitness of an organism, such a link need not exist for all TFs. Moreover, for TF's regulating many genes, the relationship between binding affinity and fitness likely varies from binding site to binding site. However, we stress that this limitation is not restricted to binding affinity, but rather applies to the quantitative phenotypes measured to date in almost all empirical adaptive landscapes of proteins and RNAs, including catalytic activity, stability, or affinity to a target ligand[3,19]. Fourth, while our measure of peak accessibility is inspired by the weak-mutation, strong-selection regime of population genetics, not all TF binding sites are under strong selection[46]. Our measure is therefore conservative, because relaxed selection for binding affinity would only make a landscape's peaks more accessible. Fifth, our landscape navigability measures assume that valleys cannot be crossed. This assumption may be violated if the mutation rate is high or if the population size is small. Valley crossing may also be facilitated by relaxed selection for TF binding, genetic hitchhiking, or by temporally changing environments[93]. A final caveat is that we only consider mutations in TF binding sites, even though mutations in the DNA binding domains of TFs are also important for regulatory evolution[80,94,95]. Data describing the effects of such mutations on a TF binding preferences are now available[80,96,97], but due to the hyper-astronomical size of protein space[27], they are far from exhaustive, capturing only a small

subset of all possible mutations to a TF's DNA binding domain, and thus precluding a comprehensive analysis.

5. Supplementary references